Universidade de Aveiro Departamento de Química
2011

João Tiago Viana de Matos

Aplicação de cromatografia líquida bidimensional a misturas complexas

Application of two-dimensional liquid chromatography to complex mixtures

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Química - Especialidade em Química Analítica e Qualidade, realizada sob a orientação científica do Doutor Armando da Costa Duarte, Professor Catedrático do Departamento de Química da Universidade de Aveiro, e da Doutora Regina Maria Brandão de Oliveira Duarte, Investigadora Auxiliar do Centro de Estudos do Ambiente e do Mar (CESAM) da Universidade de Aveiro.

Aos meus pais e irmã.

O Júri

Presidente

Prof. Doutor Artur Manuel Soares da Silva
professor Catedrático do Departamento de Química da Universidade de Aveiro

Prof. Doutora Teresa Alexandra Peixoto da Rocha Santos
professora Associada do Instituto Piaget de Viseu

Prof. Doutor Armando da Costa Duarte
professor Catedrático do Departamento de Química da Universidade de Aveiro

Doutora Regina Maria Brandão de Oliveira Duarte
investigadora Auxiliar do Centro de Estudos do Ambiente e do Mar (CESAM) da Universidade de Aveiro

Agradecimentos

A realização desta dissertação não seria possível sem a ajuda e colaboração de diversas pessoas, a quem endereço os meus mais sinceros agradecimentos, nomeadamente:

Aos meus orientadores, Doutor Armando da Costa Duarte e Doutora Regina Maria Brandão de Oliveira Duarte, pelo tempo, disponibilidade e compreensão demonstrada ao longo de todo o trabalho desta dissertação.

À minha família por todo o apoio e suporte financeiro que me deram ao longo dos meus estudos.

Aos meus colegas e amigos que sempre me apoiaram e ajudaram nos bons e maus momentos.

A todas as outras pessoas que direta ou indiretamente contribuíram para a realização deste trabalho.

A todos muito Obrigado.

Palavras-chave

Cromatografia líquida compreensiva bidimensional; Separações multidimensionais; Deteção de picos; Pureza de picos bidimensionais; Função de resposta cromatográfica bidimensional; Misturas orgânicas complexas; Impressão digital de vinhos.

Resumo

A utilização de sistemas avançados de cromatografia, nomeadamente a cromatografia bidimensional acoplada a detetores multidimensionais, permite obter grandes quantidades de dados que necessitam de especiais cuidados ao serem processados, de modo a caracterizar e quantificar, tanto quanto possível, os analitos em estudo. Esta dissertação começa por identificar as principais tendências, estudos necessários e lacunas das metodologias de processamento de dados multidimensionais obtidos em cromatografia compreensiva bidimensional. Para além de vários conceitos da cromatografia unidimensional que ainda não foram testados no modo bidimensional, verifica-se a premência para a melhoria e desenvolvimento de algoritmos e software para processamento dos dados de cromatografia compreensiva bidimensional.

Assim, este trabalho apresenta uma nova função de resposta cromatográfica bidimensional (CRF_{2D}), a qual é testada para estimar o índice de qualidade da separação de misturas orgânicas complexas por cromatografia líquida compreensiva bidimensional. Esta função baseia-se no conceito de "pureza de picos" da cromatografia líquida unidimensional, o qual foi redefinido para o modo bidimensional. A nova CRF_{2D} também inclui outros critérios de qualidade da separação, nomeadamente o número de picos bidimensionais do cromatograma e o tempo de análise. O desempenho da CRF_{2D} foi comparado com o de uma métrica de resolução já existente, usando para o efeito cromatogramas simulados. A capacidade da nova função para qualificar o grau total de separação que é obtido sob diferentes condições cromatográficas foi avaliada através de um estudo de uma mistura de quatro compostos aromáticos.

Finalmente, o conhecimento obtido no decorrer desta dissertação foi usado para desenvolver um protocolo para a análise de três amostras portuguesas de vinho tinto, usando um sistema de cromatografia líquida compreensiva bidimensional acoplado a um detetor de fotodíodos e a um detetor de fluorescência, de forma a obter a respectiva impressão digital cromatográfica bidimensional. As impressões digitais foram quantificadas usando a CRF_{2D}, cujos resultados foram utilizados para comparar as amostras de vinho tinto.

Keywords

Comprehensive two-dimensional liquid chromatography; Multidimensional separations; Peak detection; Two-dimensional peak purity; Two-dimensional chromatographic response function; Complex organic mixtures; Wine fingerprinting.

Abstract

The operation of advanced chromatographic systems, namely comprehensive two-dimensional chromatography coupled to multidimensional detectors, allows achieving a great deal of data that need special care to be processed in order to characterize and quantify as much as possible the analytes under study. This dissertation starts by identifying the main trends, research needs and gaps on the techniques for data processing of multidimensional data sets obtained from comprehensive two-dimensional chromatography. Apart from using other concepts from one-dimensional chromatography, which have not been tested for two-dimensional chromatography, there is still room for new improvements and developments in algorithms and software for dealing with two-dimensional comprehensive chromatographic data.

Therefore, this work presents a new two-dimensional chromatographic response function (CRF_{2D}) for the estimation of the quality index of separation in comprehensive two-dimensional liquid chromatography of complex organic mixtures. This objective function is based on the concept of peak purity for one-dimensional liquid chromatography, which has been redefined for comprehensive two-dimensional liquid chromatography. The new CRF_{2D} also includes other separation quality criteria, namely the number of two-dimensional peaks appearing in the chromatogram and the analysis time. The performance of the developed CRF_{2D} was compared to an existing resolution metric, using simulated chromatograms. The capability of the new function to qualify the overall separation degree that it is attained under different chromatographic conditions was further assessed through a comprehensive two-dimensional liquid chromatography study of a mixture of four aromatic compounds.

Finally, the knowledge obtained along this dissertation was used to develop a procedure to analyse three Portuguese red wine samples by comprehensive two-dimensional liquid chromatography coupled to a diode array and a fluorescence detectors, in order to obtain a two-dimensional chromatographic fingerprinting. The overall quality of chromatographic separation was quantified using the CRF_{2D}, and the red wine samples were compared taking into account the obtained CRF_{2D} values.

Table of contents

List of figures

List of tables

List of abbreviations

ALS	Alternating least squares
APTLD	Alternating penalty trilinear decomposition
ATLD	Alternating trilinear decomposition
AU	Arbitrary unit(s)
CE	Capillary electrophoresis
CRF	Chromatographic response function
CRF_{2D}	Two-dimensional chromatographic response function
DAD	Diode array detector
DOC	Controlled denomination of origin
DTLD	Direct trilinear decomposition
FLD	Fluorescence detector
GC	One-dimensional gas chromatography
GC × GC	Comprehensive two-dimensional gas chromatography
GRAM	Generalized rank annihilation method
HPLC	High-performance liquid chromatography
IKSFA	Iterative key set factor analysis
LC	One-dimensional liquid chromatography
LC × LC	Comprehensive two-dimensional liquid chromatography
MS	Mass spectrometer
EEM	Excitation-emission matrix.
MCR	Multivariate curve resolution
PARAFAC	Parallel factor analysis
SWATLD	Self-weighted alternating trilinear decomposition
TOFMS	Time-of-flight mass spectrometry
WTTFA	Window target testing factor analysis
%RSD	Percentage of relative standard deviation

I

Aims and structure of dissertation

1.1. Introduction

Nowadays, the increasing need for dealing with complex mixtures from different science fields, (i.e. genomic, proteomic and metabolomics), and the lower potential of one-dimensional separation techniques to differentiate complex mixtures has raised the urgency for developing new separation methods, mostly based on comprehensive two-dimensional chromatographic techniques. This demand is supported by the increasing number of publications and citations in the last 12 years focusing on this issue, as shown in Fig. I-1.

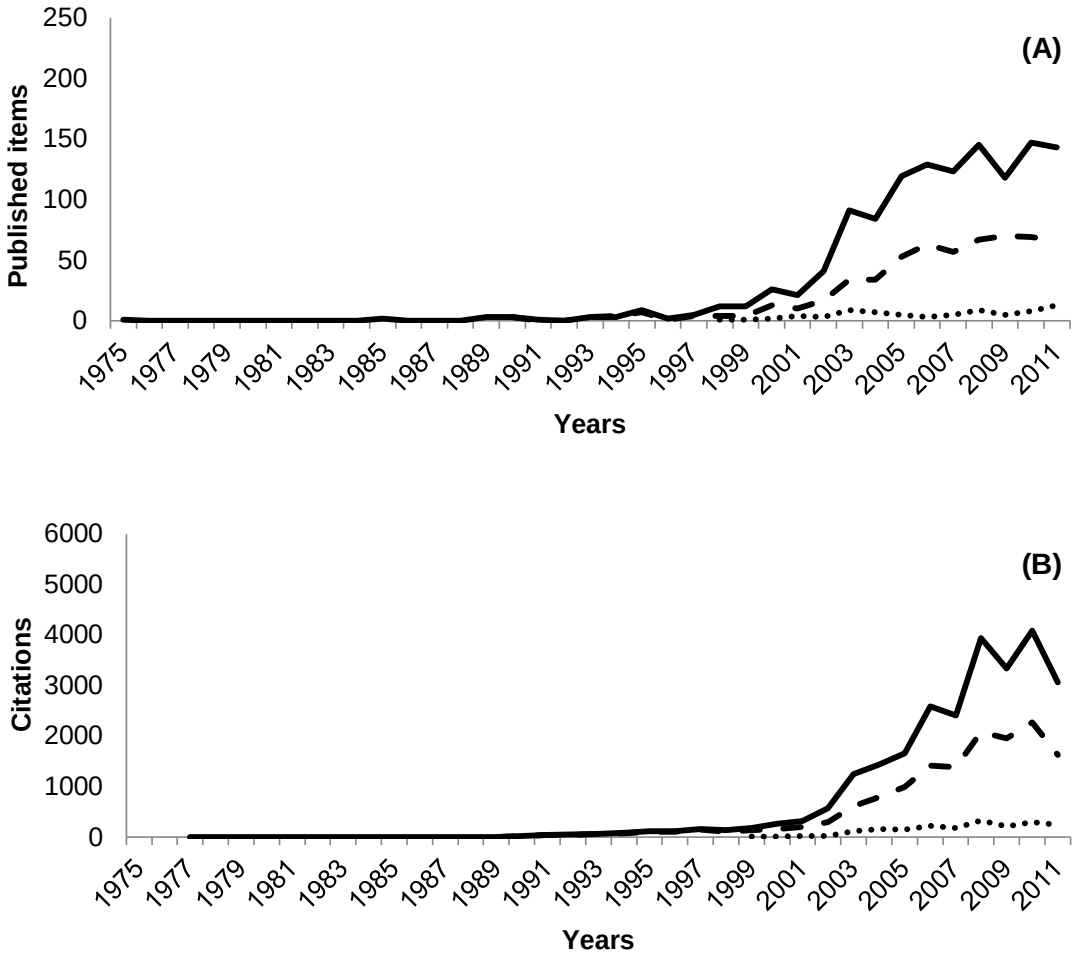

Fig. I-1. Number of published items in each year **(A)** and citations in each year **(B)** for the search terms of *"Comprehensive two-dimensional* chromatography" (solid line), "Comprehensive *two-dimensional liquid* chromatography" (dashed line), and "*Comprehensive two-dimensional liquid chromatography* chemometrics" (dotted line). The search has been made in all databases of the science search engine *ISI web of knowledge* (*http://apps.isiknowledge.com, 6-Oct-11).*

For the purposes of this research work, and after refining the search results for the topic "comprehensive two-dimensional liquid chromatography" (LC × LC), it is possible to verify that this technique follows a trend similar to that shown in Fig. I-1, with the prospect of further evolving in the upcoming years. Nevertheless, the current volume of publications using this technique can still be considered relatively low taking into account its own potential. On the other hand, the major drawback normally associated with the application of these techniques is the handling and processing of the multi-dimensional chromatographic data. Usually, the type of data obtained and, consequently, all data treatment, including algorithms and methodologies developed for this purpose, are common to most of the comprehensive two-dimensional techniques (in some cases with minor modifications). However, the amount of work performed in this field is still insufficient and in most cases the published works are too complex for most users. Consequently, much work is still needed in this field of research.

1.2. Aim of the work

Given the gaps that still exist in the field of data processing for LC × LC, specially for dealing with complex mixtures, one of the main goal of this research work is to make an attempt to address this problem. Therefore, the first specific objective of this dissertation was to review the studies that have been developed in this area and to identify some of the limitations of the existing tools for data processing.

Another specific objective was to develop all the concepts and tools necessary to build a new chromatographic response function (CRF) for assessing the separation quality in LC × LC, and to demonstrate the application of this new mathematical model in simulated and real two-dimensional chromatograms.

Finally, the third specific objective entailed the optimization of the chromatographic conditions of an LC × LC system and, alongside the knowledge acquired in the previous tasks, develops a methodology for fingerprinting Portuguese red wine samples.

The work of this dissertation originated two scientific papers, one entitled “A new chromatographic response function for assessing the separation quality in comprehensive two-dimensional liquid chromatography”, which was accepted in the *Journal of Chromatography A*, and another one entitled “Trends in data processing of comprehensive two-dimensional chromatography: state of the art”, which was submitted to a special issue in “Chemometrics in Chromatography” on *Journal of Chromatography B*.

1.3. Dissertation structure

This dissertation is organized into five chapters, starting with a general description of the dissertation and a contextualization of the scientific relevance and specific objectives of the research work (Chapter I). Chapter II reviews the trends in data processing for application into the different two-dimensional chromatographic techniques, including data acquisition and handling, peak detection and quantification, measurement of overlapping of two-dimensional peaks, and data analysis software. In Chapter III, a new chromatographic response function is proposed aiming at evaluating the quality of chromatographic separation in two-dimensional chromatography, followed by its application into the LC × LC analysis of complex organic mixtures. This third chapter also revisits the existing concepts for peak detection and measurement of resolution, and presents new approaches to qualify the degree of chromatographic separation. Based on the concepts previously developed, Chapter IV presents a procedure for the LC × LC analysis of three Portuguese red wine samples and further comparison of their chromatographic fingerprint. Finally, Chapter V presents the main conclusions drawn from the current study and suggestions for future research.

II

Trends in data processing of comprehensive two-dimensional chromatography

2.1. Introduction

The development of several one-dimensional separation techniques, such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE), led to the generalized idea by the end of the twentieth century, that these techniques could be just finely tuned in order to solve all the practical problems in Analytical Chemistry (Stoll *et al.*, 2007). However, the need for analysis of increasingly complex samples with a large number of compounds, highlighted the limitations of such techniques, and prompted the development of technologies with a much higher separation capacity in order they could take full advantage of coupling them to advanced detection systems, such as mass spectrometry and nuclear magnetic resonance spectroscopy.

The need for improving the analytical figures of merit associated to the research explosion in proteomics and metabolomics, and the ever increasing requirements for adequate identification and quantification of proteins, glycoproteins and metabolite products has prompted a need to push separation techniques to their limits. Furthermore, even when one-dimensional chromatography could produce acceptable results, they do not have the separation power to deal with complex samples, and their use in such cases would mean spending a lot of time for analysis (Guiochon *et al.*, 2008, Stoll *et al.*, 2007). The obvious response to this lack of separation power of one-dimensional techniques is the development of multidimensional chromatographic systems using two or more independent separation mechanisms.

Multidimensional separation can be understood as a separation system capable of discriminating the components from a mixture, using different separation mechanisms which are connected but do not interact among themselves, that is, they should be completely independent from each other. There are two modes of operation of multidimensional chromatography: heart-cutting and comprehensive. In the heart-cutting mode, only some selected fractions are transferred from the first into the second separation system, and the results become two separate one-dimensional data sets. On the other hand, a multidimensional chromatography is comprehensive when all fractions separated in the first separation system are successively injected into further separation systems,

comprising a fully automated methodology and providing complete chemical information of the whole sample in the different chromatographic dimensions.

As reviewed by Phillips and Beens (1999), the comprehensive multidimensional chromatography became more relevant after the development of comprehensive two-dimensional gas chromatography (GC × GC), more than a decade after the development of a first LC × LC by Erni and Frei (1978). Although the data sets resulting from GC × GC have received more attention than those resulting from any other comprehensive two-dimensional chromatographic technique, they are formally equivalent and most of the work developed for GC × GC can be applied with small modifications to other chromatographic combinations, such as LC × LC, LC × GC, and LC × CE.

The data collected from advanced chromatographic systems designed for the analysis of complex samples contain huge amounts of information that need complex processing algorithms in order to take advantage of such powerful analytical systems. For instance, analysis of a sample with *n* replicates in a two-dimensional chromatographic system coupled to a multichannel detector, such as a diode array detector (DAD) or a mass spectrometer (MS) can produce quadrilinear data sets. Fig. II-1 shows how dimensionality of data sets may increase with increasing analytical dimensions: the data structure starts as one-way (Fig. II-1A) associated to one-dimensional chromatography coupled to a single channel detector, and it evolves to a two-way structure associated to a two-dimensional chromatography coupled to a single channel detector, then reaches a three-way structure (Fig. II-1B) associated to a two-dimensional chromatography coupled to a multichannel channel detector, and finally attains a four-way data structure (Fig. II-1C) obtained from *n* replicates in a two-dimensional chromatography system coupled to a multichannel detector.

In this chapter it is intended to discuss the state of the art in data processing for multidimensional data sets obtained in different types of two-dimensional chromatography, from the pre-treatment until the quantification of the identified chromatographic peaks. The discussion will lead to the identification of the main trends in data processing of comprehensive two-dimensional chromatography and it will pinpoint the gaps and research needs that should be tackled in this field.

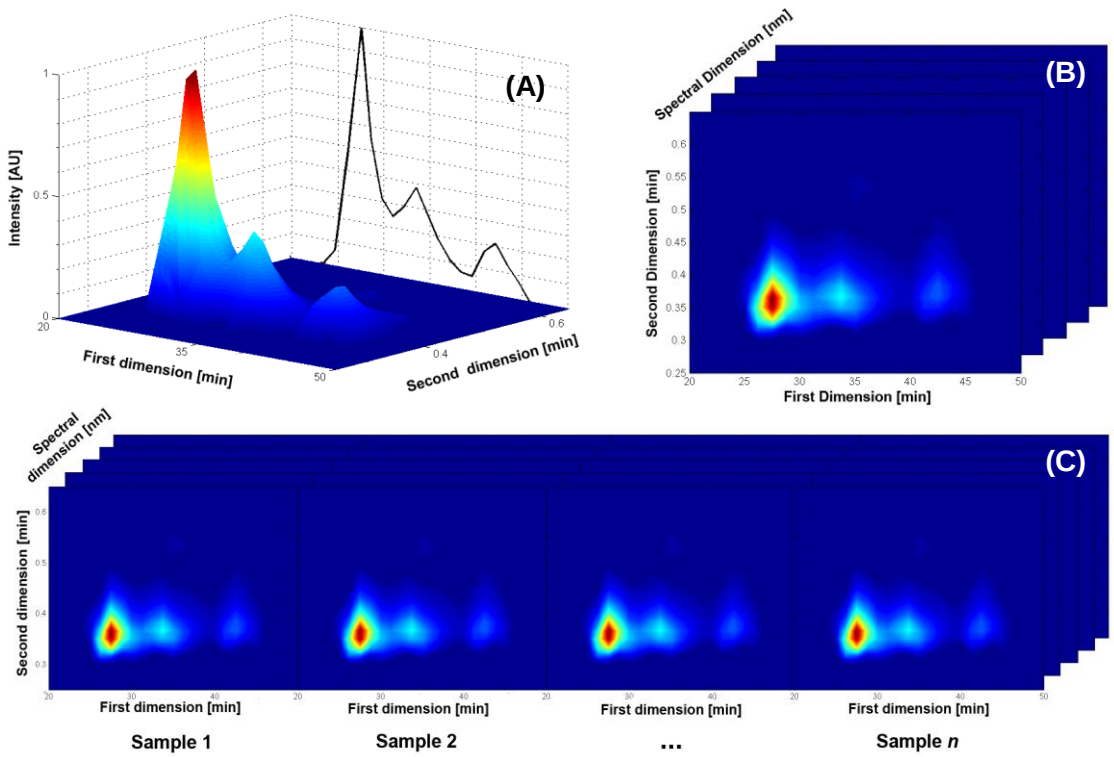

Fig. II-1. Representation of (**A**) one-dimensional chromatography (one-way data) two-dimensional chromatography (two-way data); (**B**) two-dimensional chromatography coupled with a multichannel detector (three-way data); and, (**C**) two-dimensional chromatography coupled with a multichannel detector with sample replicates (four-way data).

2.2. Data acquisition and handling in comprehensive two-dimensional chromatography

The massive amount of data generated from the current high-resolution analytical instrumentation requires the use of computerized assistance for data processing and transformation. The two-dimensional chromatography is no exception, and the use of informatics tools has become essential for transforming the raw analytical data into fit for purpose information. The two-dimensional chromatography produces a considerable amount of data in a relatively short time when applied to the separation of complex mixtures. Such an enhancement in performance provides an order-of-magnitude increase in peak separation capacity, when compared to one-dimensional chromatography.

Handling of two-dimensional chromatography data is a challenging task namely in Analytical Chemistry. The acquisition of data in real time from two-dimensional chromatography coupled to detectors, such as DAD or MS, generates huge data files, that can reach more than 10 million data points which may lead to considerable problems in storage and processing (Guiochon *et al.*, 2008, Porter *et al.*, 2006). The greatest challenges lies in producing automatic tools capable of processing and converting the data matrix under useful forms without losing control on the analysis of samples for obtaining raw data, and transformation of data into useful chemical information. The scarcity of software available for data acquisition and handling in two-dimensional chromatography, especially in LC × LC, is one of the most significant impediments for a generalized adoption of these separation techniques.

There are three general approaches that can be used to deal with the two-dimensional chromatographic data. The first approach, and since the two-dimensional chromatograms can be considered as a set of consecutive one-dimensional chromatograms, is to deal with those chromatograms individually using all the data treatment tools already extensively developed for one-dimensional chromatography; the second approach is to deal directly with the two or higher order dimensional matrix after a previous step of data modulation; and finally, the third approach is to transform the data in an image file (also after a data modulation step) and follow further image treatment and processing. The first approach can take advantage of a full body of knowledge already designed and extensively studied for one-dimensional chromatography and the availability of a large amount of software ready for use. However, considering the long-time of analysis compared to the very short time necessary for modelling, there are a large number of one-dimensional chromatograms, and consequently there is too much time spent on each individual chromatogram. On the other hand, when dealing with two or higher order data matrix, as it is the case of the second approach, it requires knowledge and expertise on complex chemometric algorithms. This is a developing field, and the evolving work is very promising, especially for dealing with three or higher order dimensional data. Finally, and despite the use of imaging tools to deal with two-dimensional chromatograms may look like a misfit, this third approach has been in fact translated into commercially available software showing excellent results (Reichenbach *et al.*, 2003, 2004, 2005, 2009)

2.2.1. Data pre-treatment

In a typical experiment of two-dimensional chromatography, the massive amounts of data acquired in both dimensions create large data files, which need to be processed in order to become useful. These data sets are extremely important in chemical analysis, and require the use of computational systems to process and extract the maximum chemical information possible. However, this processing requires different approaches and methodologies depending on the different ends or goals set by the analyst. Therefore, in order to accomplish this task, it becomes necessary a transformation of the data into a more suitable form, in accordance with the aim of the analytical work (Guiochon *et al.*, 2008).

In general terms, the most common operations and important methodologies in data processing and pre-treatment in order to become possible the representation, detection and quantification of peaks are, the removal and correction of background, the attenuation of the signal noise, the correction of uncontrolled shifts in retention time, the identification and removal of signal artifacts, and the resolution of overlapping peaks (Guiochon *et al.*, 2008, Reichenbach *et al.*, 2004, Stoll *et al.*, 2007).

2.2.1.1. Modulation and interpolation of data

The first step in data processing of comprehensive two-dimensional chromatography is to extract the experimental data from the detector response and build the corresponding two-dimensional chromatogram. In the first applications of two-dimensional separations, namely the development of LC × LC by Erni and Frei (1978), two detectors were used for measuring the analytical signals at the end of each chromatographic column. Nowadays, due to both the comprehensive methodology of the analysis and the available software, the experimental signal is only measured at the end of the second column. Consequently, in order to transform the output of the detector positioned at the end of the second column into a two-dimensional chromatogram, it is necessary to slice the output according to the modulation time of the switching valve, and

re-organize each sliced chromatogram along the time axis of the volume separated in the first column in this same modulation time.

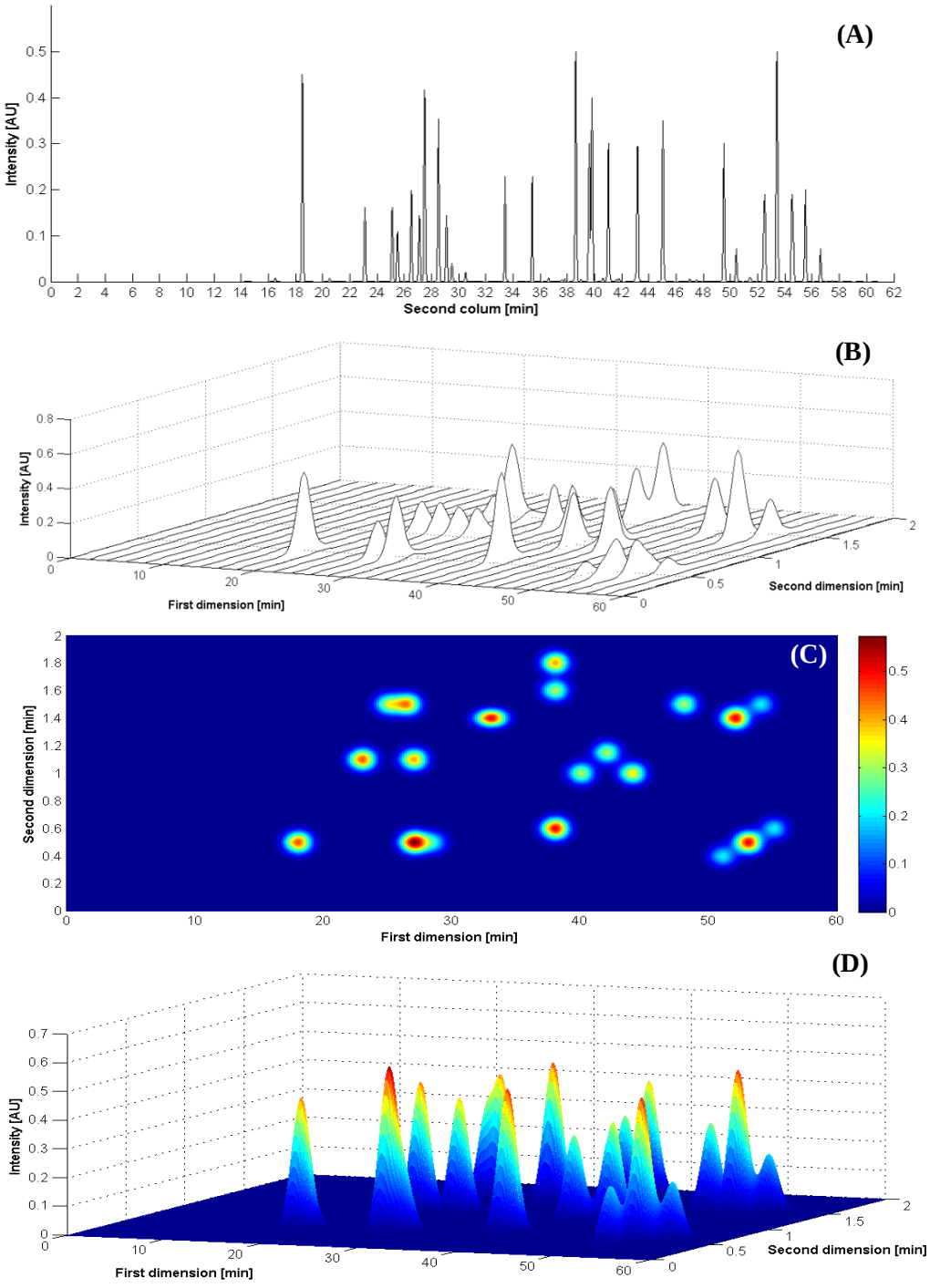

Fig. II-2. Representation of a simulated data set for LC × LC: (**A**) raw data from the detector positioned at the end of the second column; (**B**) layout of the sliced one-dimensional chromatograms defined by the modulation time; (**C**) contour plot of a two-dimensional chromatogram after smoothing; and, (**D**) three dimensional plot of a two-dimensional chromatogram after smoothing.

Fig. II-2A shows a simulated output from detector positioned at the end of a two-dimensional separation system, with a modulation time of 2 minutes. Since the modulation time is known, then the application of the slicing methodology above mentioned is straightforward, and the resulting Fig. II-2B shows the same data set but now under a form of a two-dimensional chromatogram. After data modulation, it may be necessary to smooth the chromatogram by interpolation of data points between one-dimensional chromatograms mainly along the first dimension in order to obtain a two-dimensional chromatogram as shown in Fig. II-2C and Fig. II-2D.

Nowadays, the availability of multichannel detectors coupled to chromatographic systems allows obtaining a range of spectral data instead of a single intensity reading, and therefore the operator can have an additional dimension in the data array related to the spectral information. Prior to the slicing of the chromatograms according to the modulation time, the operator faces already a two-way data array that should be taken into account. A two-way example of this experimental data, acquired in a LC × LC-DAD, is represented in Fig. II-3, with an experimental modulation time of 1.25 min and wavelength information from 215 to 300nm. Therefore, to convert the experimental data into usable chemical information it is necessary, for each wavelength signal, to divide the data according with the modulation time for building a three-dimensional data matrix. Recently, Allen and Rutan (2011) compared five different interpolation methods, including linear interpolation followed by cross correlation, piecewise cubic Hermite interpolating polynomial, cubic spline, Fourier zero-filling, and Gaussian fitting, and the authors concluded that these methods performed equally well.

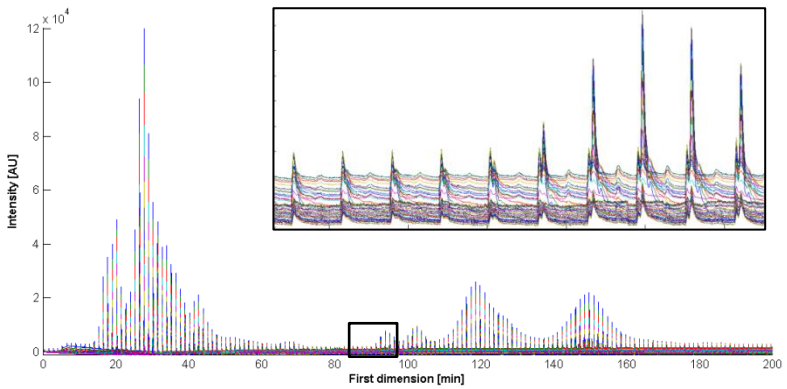

Fig. II-3. Representation of a simulated data set for LC × LC-DAD chromatographic data with expansion of a section between 84 and 98min.

2.2.1.2. Data representation and visual features

For representation of a two-dimensional data matrix either the contour colour coded plot or the three-dimensional plot are the types of graphs that easily allow visualizing and understanding the distribution of peaks in a chromatogram. As shown in Fig. II-2C, such a graph does not provide quantitative information but it is an excellent visualization tool which becomes the first essential step for further processing of the data obtained. The use of a colour coded scale for peak intensity allows an immediate qualitative assessment of the different peaks and respective heights. Fig. II-2D uses the same data as in Fig. II-2C for simulating a typical two-dimensional chromatogram, and the three-dimensional plot provides already quantitative information, thus becoming the best approach to observe, analyse, and comment the visual features of three-dimensional peaks shape and form (Guiochon *et al.*, 2008, Reichenbach *et al.*, 2011). When dealing with higher order data dimensions, it is not possible to produce conventional contours and three-dimensional plots, since a two-dimensional chromatographic system coupled to a multichannel detector will produce a four dimensional data array.

2.2.1.3. Background and noise signal

As in any other analytical methodology (Amigo *et al.*, 2010), it is possible to separate the two-dimensional chromatographic experimental data signal (Fig. II-4A) in to three major parts: the analytical signal, the background signal, and the noise. The analytical signal, as depicted in Fig. II-4B, gives the chromatographic response of the analyte obtained for a particular set of experimental conditions and a particular detector free off any traces of noise and background effects. The background, represented in Fig. II-4C, is associated with the systematic response from the chromatographic system not related to the analyte. Finally, the noise, represented in Fig. II-4D, is associated with random variations and it is usually related to the sensitivity of the detector.

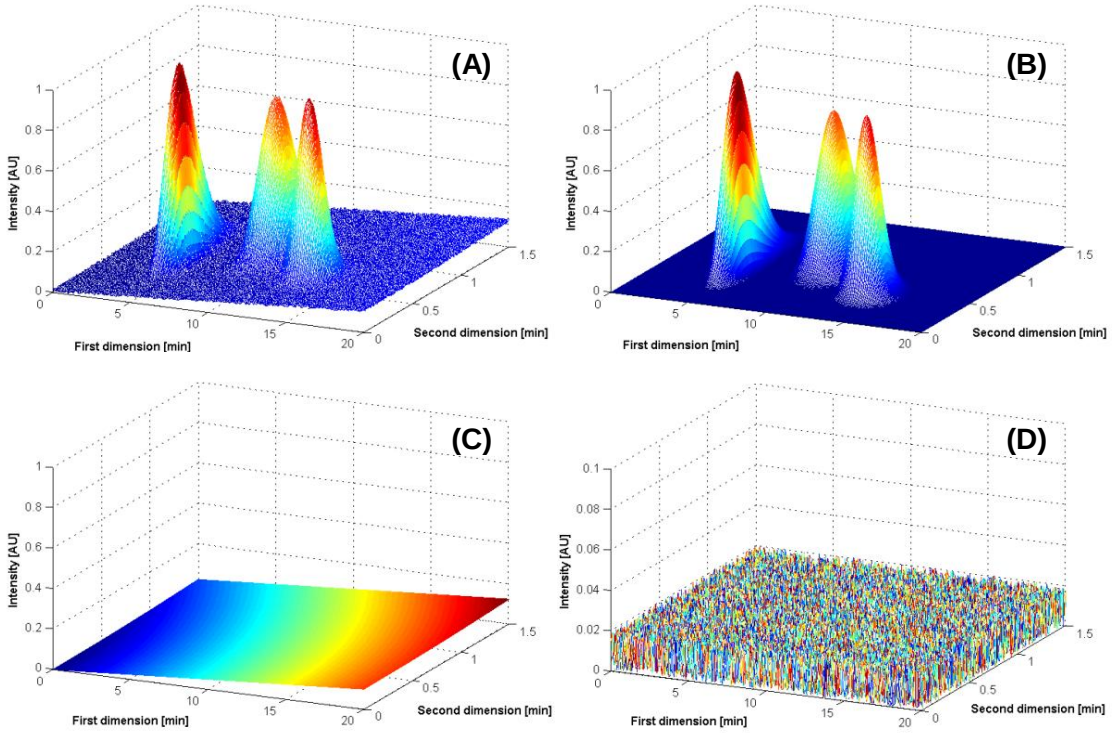

Fig. II-4. Representation of a simulated 2D chromatogram and its component parts (**A**): analytical signal (**B**), background (**C**) and noise (**D**) (visualization inspired by Amigo *et al.* (2010)]).

The background signal and the noise can cause many problems, particularly for quantification, since they can change both the shape and the elution time of peaks. This background drift can be caused by changes in the composition of mobile phase when in gradient mode, variation of pressure and temperature, and fluctuations caused by the injection valve. When the signal-to-noise ratio is low, it becomes difficult in practice to separate between noise and background, but in order to achieve a clear analytical signal with a flat baseline, it is necessary to identify and eliminate the interferences caused by noise and background signal. Although the noise interference can be usually reduced using smoothing algorithms, the strong background drifts are either difficult to resolve. One of the easiest ways to deal with the background interference is performing a "mean centering" of the data, but only in the presence of a relatively stable background. Another possible way is to subtract a “blank” chromatographic run obtained in the same instrumental conditions, but even so this strategy may not be very accurate due to possible changes in eluent spectrum (Zhang *et al.*, 2007) and differences in noise between chromatograms. For overcoming the drawbacks of using simple strategies in two-dimensional chromatography,

more complex algorithms have been proposed for removing the background signal and noise.

Zeng *et al.* (2011) apply to each one-dimensional peak of a two-dimensional chromatogram a baseline correction and a moving windows average method for data smoothing. On the other hand, Zhang *et al.* (2007) suggested the use of the trilinear decomposition method to remove the three-dimensional background drift in two-dimensional chromatography coupled with multichannel detectors (in this case a LC × LC coupled to a DAD system). The authors (Zhang *et al.*, 2007) used the background drift and the analytical signal as factors for building a factor model by applying an alternating trilinear decomposition (ATLD) algorithm to the raw dataset in order to extract the background factor, and then subtract it from the raw dataset. The ATLD algorithm showed adequate properties of convergence and robustness to the excess of factors used, although others chemometric methods such as parallel factor analysis (PARAFAC), self-weighted alternating trilinear decomposition (SWATLD) and alternating penalty trilinear decomposition (APTLD) algorithms can be used with similar results (Zhang *et al.*, 2007). Finally, this methodology removes the background drift without the loss of peak information in entire spectral region and without the need for replicates and "blank" chromatograms.

An alternative technique is the use of image treatment software to estimate and remove the background from images files of two-dimensional chromatograms, as suggested and applied by Reichenbach *et al.* (2003) for GC × GC. This algorithm takes advantage of the following structural and statistical properties of the background from the images of two-dimensional chromatograms: dead-bands that are regions without analytical signal; the mean of background level does not change much when compared to the characteristic peak widths; and, the noise present has the same statistical properties of the random white noise. This background-removal algorithm has been included in the GC Image and LC Image software packages, which has been reported in several works (Reichenbach *et al.*, 2003, 2004, 2005, 2009).

2.2.1.4. Correction of shifts in retention time of peaks

One of the most crucial steps in any chromatography, either one-dimensional or multidimensional, is to ensure the precision in the determination of retention time for each and every peak. This run to run time deviations are often observed in the chromatographic analysis, and can be easily identified by the comparison to the patterns of variation between replicates or standards. Due to the high sensitivity of the analytical methods, a poor precision of the retention times, when not corrected, can generate enormous deviations in the most of the chromatographic detection and quantification techniques actually used.

The fluctuations in retention time of peaks are always present in two-dimensional chromatographic systems and can be originated from variations in temperature and pressure, degradation of the stationary phase, and matrix effects. For data processing to be sucessful, there is a need to ensure that the retention times between replicates are repeatable and reproducible, the time axes are synchronized, and the peaks are properly aligned. This alignment or warping can be accomplished by using different algorithms. Fraga *et al.* (2001) suggested an alignment technique of the retention times based on minimizing residuals in the generalized rank annihilation method GRAM which was addressed by Prazen *et al.* (1998) for second-order hyphenated chromatography, namely GC-MS. On the other hand, van Mispelaar *et al.* (2003) suggested a correlation-optimized shifting, based in an inner-product correlation associated with selected regions of the GC × GC data. This algorithm uses a two-dimensional chromatogram reference to align all selections and as this alignment is performed, the inner-product correlation is calculated in order to identify the best-fit position. Johnson *et al.* (2004) described an algorithm based on windowed rank minimization alignment with interpolative stretching between the windows. In this work (Johnson *et al.*, 2004) the algorithm was used to deal with shifted GC × GC retention times during the quantification of naphthalene in jet fuel, and it produced better results than in the case where quantification of the chromatograms was not pre-aligned. Pierce *et al.* (2005) reported the application of a comprehensive two-dimensional retention time alignment algorithm that allows a warping in both chromatographic dimensions using a novel indexing scheme, and preserves the separation

information in both dimensions. The algorithm was applied by Pierce *et al.* (2005) to GC × GC but it can be applied to any two-dimensional separation system with a gain on the retention time precision and also restoring the trilinearity of the data without losing quantitative information. Zhang *et al.* (2008) developed a two-dimensional Correlation Optimized Warping Algorithm, (2D-COW) to align data obtained from two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC × GC-TOFMS). This powerful and flexible algorithm stretches and compresses a local sample segment of the two-dimensional chromatogram in order to maximize the correlation from the sample relatively to the two-dimensional chromatogram reference. Such calculations allow interpolating the warp non-grid points from the shifted grid points in order to align the chromatograms. When using image based software, the method for automatically aligning chromatograms developed by Hollingsworth *et al.* (2006) can be applied, as in the case of Nelson *et al.* (2006) and Wardlaw *et al.* (2008) to study the weathering of an oil spill and oil seepage, and Cordero *et al.* (2010) to compare coffee samples. Unfortunately, all these methodologies and algorithms cannot deal with three- or higher-way data sets, as those obtained from two-dimensional chromatography coupled to multichannel detectors. As the degree of hyphenation of several detectors to two-dimensional chromatography increases, the more urgent becomes the search for more sophisticated techniques for correction of shifts in retention times of peaks. Recently, Allen and Rutan (2011) developed an algorithm especially suited to LC × LC-DAD that allow to deal with four-way data with satisfactory results. Another issue that it is necessary to take into account in two-dimensional chromatography is the different time scale between the second dimension and the modulation period. Since the two-dimensional chromatogram is generated from a one-dimensional signal array, it is fundamental to make sure that there are not two-dimensional peaks "wrapped-around" (Micyus *et al.*, 2005), i.e., the maximum range of a peak retention time in the second dimension is lower than the modulation period. One way to deal with this problem, which is quite common in chromatography of complex mixtures, is an algorithm capable of finding the absolute retention times in these cases as suggested by Micyus *et al.* (2005). In this algorithm, after the detection of "wrapped-around" two-dimensional peaks, a series of chromatograms are re-analysed by an integer fraction of original modulation period and shifts of the retention of the second dimension are used to determine the absolute retention times.

2.3. Peak detection in comprehensive two-dimensional chromatography

After data pre-treatment, the most critical step in the analysis of two-dimensional chromatograms is probably the detection of a two-dimensional peak. In the past decade, many authors dedicated to develop algorithms capable of detecting two-dimensional peaks, especially for applications in GC × GC. A large part of these methodologies are a direct consequence of the already developed algorithms for one-dimensional chromatography since, in first instance, the two-dimensional separations can be understood as an upgrading of the classical column chromatography by just adding another column as a second dimension. However, with the evolution of the concept of two-dimensional chromatography, associated to the increasing use of multichannel detectors and the consequent production of a huge amount of data, it is foreseeable an augment of methodologies and algorithms that can deal with these data more efficiently in a global manner, rather than adding up one-dimensional chromatograms.

Conceptually, the peak detection algorithms can have two main purposes: in non-targeted analysis, the algorithms should be capable of identifying two-dimensional peaks in complex mixtures without any pre-information from de sample, such as the two-step peak detection algorithm (Peters *et al.*, 2007a) and the watershed algorithm (Reichenbach *et al.*, 2004); in targeted analysis, the algorithms need a complete or a regional two-dimensional chromatographic reference, and the peak detection is made by comparison with this standard, such as the GRAM method (Sanchez and Kowalski, 1986). Several authors, who have developed algorithms for two-dimensional peak detection, have also developed methods for quantification, since most of the analytical work involves not only the screening of peaks but also their quantification. Finally, it is important a word of caution regarding the choice of an algorithm because it may have to take into account several factors such as speed of analysis, accuracy of results, and components of interest, depending on the specific work involved.

2.3.1. Two-step peak detection algorithm

One of the most important contributions to the actual state of the art in two-dimensional chromatography has been produced by Peters *et al.* (2007a), namely when suggesting new methodologies and important concepts for detection and resolution of peaks in GC × GC. Theoretically, the methodology can also be applied to LC × LC with minor modifications, if any at all. For the specific case of peak detection, Peters *et al.* (2007a) developed an algorithm that deals with this problem in two main steps: firstly, already known methods developed for peak detection in one-dimensional chromatography are applied to detect the peaks one dimension at a time; secondly, some criteria are used to decide which peaks of the first dimension correspond to the same compound in the second dimension, that is, some criteria are set up for the process of merging the peaks from the first dimension with the peaks of the second dimension that are all produced by the same compound .

The raw data acquired in comprehensive two-dimensional chromatography is constituted by a set of one-dimensional chromatograms, each corresponding to a single injection in the second column. Once the obtained data has been organized in a two-dimensional matrix, it becomes possible to analyse all chromatograms using one-dimensional peak detection techniques. The two-step algorithm (Peters *et al.*, 2007a) detects one-dimensional peaks based on the properties of the derivatives of the peaks computed by the Savitsky-Golay (1964) method. From the original chromatogram and from the first- and second-order derivatives, it is possible to characterize the properties of chromatographic peaks, namely the height, and the peak starting-point and end-point (peak region). As shown in Fig. II-5A, the starting-point and the end-point of a one-dimensional peak correspond to the first and last point above zero, respectively, in the first-order derivative (Fig. II-5B). Although, in practice the first derivative often does not reach zero, such a problem can be overcome by defining a minimal value to quantify (Peters *et al.*, 2007a). On the other hand, when the value of zero for the first-order derivative coincides with the minimum of the second-order derivative (Fig. II-5C), then it means that the original chromatographic peak has reached the maximum value.

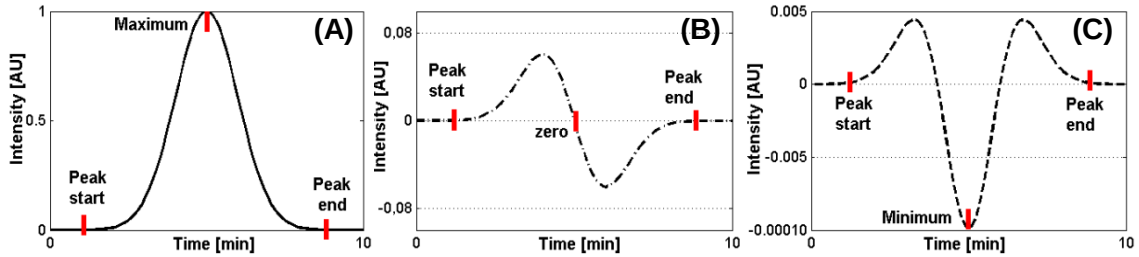

Fig. II-5. Properties of a chromatographic peak assumed to be Gaussian (**A**), its first-order derivative (**B**), and its second-order derivative (**C**).

After defining the peaks in the previous step, it is possible to apply an algorithm following some pre-defined criteria, in order to merge the one-dimensional peaks in the two dimensions. This algorithm produces a cluster with a collection of one-dimensional peaks in the consecutive chromatograms that belong to the peak of the same compound (Peters *et al.*, 2007a). Ideally, the same peak in consecutive chromatograms will elute always at the same time, but sometimes such synchronization may not happen due to several operational variables, namely in GC × GC (Peters *et al.*, 2007a) such as control and timing of cooling and heating programmes, non-linearity of distribution of isotherms, and non-instantaneous re-injections in to the second column. The process of peak merging starts off with the first peak appearing in the first chromatogram obtained in the second-dimension and the attempt of finding all the merging combinations with all the peaks found in the subsequent chromatograms of the second-dimension. Then, the overlap criterion and the unimodality criterion are applied to test whether it is possible to merge these peaks. Finally, the procedure repeats itself until all the merging combinations are tested (Peters *et al.*, 2007a).

For checking whether there is any overlap of the peak region (where the peak starts and ends) in a chromatogram, the overlap criterion examines the degree of overlap of two peak regions, one from the existent two-dimensional cluster (peak A in Fig. II-6) and the other from the candidate to be merged (peak B in Fig. II-6). The ratio of overlap, in percentage, is calculated according to the following equation, as suggested by Peters *et al.* (2007a):

$$OV = \frac{b}{a} \times 100\% \qquad \textbf{(1)}$$

where *b* is the length of the region where the two peaks overlap and *a* is the length of the peak region of peak A (Fig. II-6).

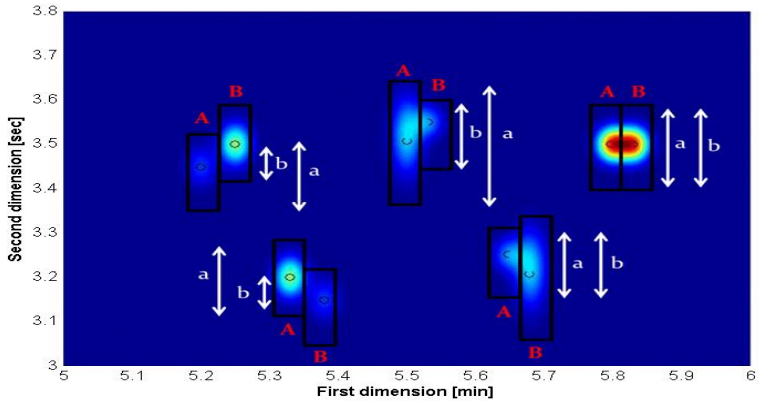

Fig. II-6. Schematic representation of peak regions of adjacent 1D peaks for different scenarios in a chromatographic map (visualization inspired by Peters *et al.* (2007a)).

After that, it is necessary to define a threshold for the acceptability of the candidate peak. If the ratio of overlap is greater than this threshold than the peak can be subjected to the next criterion; otherwise, the peak cannot be merged and the algorithm continues to the next candidate (Peters *et al.*, 2007a).

The basis for the application of unimodality criterion, as suggested by Peters *et al.* (2007a), is the analysis of the peak-maxima profile in the first chromatographic dimension, between the peaks checked for merging in two-dimensional clusters. As an example, Fig. II-7A shows the peak maxima profile for the two-dimensional chromatogram shown in Fig. II-7B. The observation of Fig. II-7A, allows concluding that there are two maxima, peak **B** and peak **D**, since each two-dimensional cluster can show only one maximum.

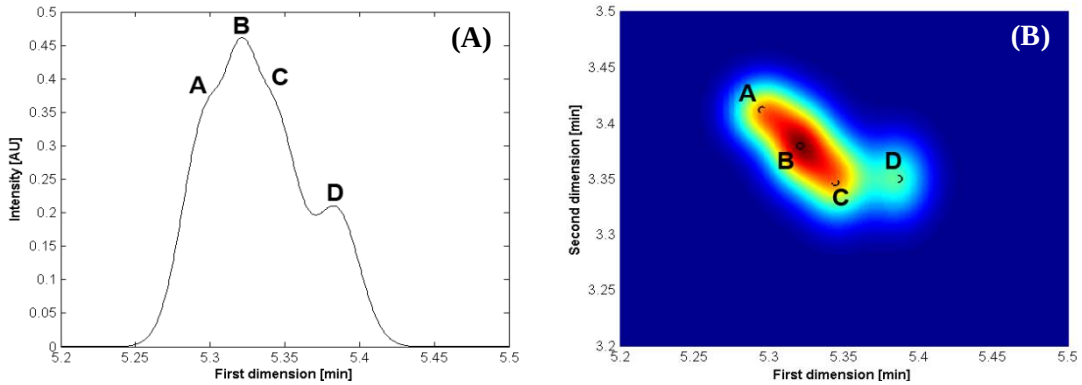

Fig. II-7. Representation of a peak-maxima profile (**A**) and its corresponding two-dimensional chromatogram after merging (**B**) (visualization inspired by Peters *et al.* (2007a)).

Finally, it is necessary to take into special consideration the handling of the closest maxima, that is, when more than one peak has been identified at the same retention time in first dimension. In such case, it becomes necessary to compare the retention time of the candidates of the second dimension, and the candidate peak with the retention time closest to the last peak of the two-dimensional cluster is then the peak to be merged.

After the application of the two-step algorithm, Peters *et al.* (2007a) also describe an integration tool for the quantification of target two-dimensional peaks: the one-dimensional peaks that were used to build the target two-dimensional peak, are integrated using a trapezoidal method and then summed in order to achieve the quantification of target analytes. The methodology based on summation of second dimension chromatograms has been also used by Pól *et al.* (2006) and Kivilompolo and Hyötyläinen (2007) for the quantification of LC × LC data. A slightly different approach based on peak volume calculation for two-dimensional contour plots has been suggested by Kivilompolo *et al.* (2008).

In order to overcome the problem related to overlapped peaks, Peters *et al.* (2007a) also suggested two alternatives: a) to integrate the area under the curve from the peak start to a perpendicular line that splits the overlapped peaks in the valley point; and b) to subtract the area under the line connecting the peak start to the valley point from the total area under the curve, as above mentioned. This method is easy to implement, and does not require any user input, which makes it ideal for quantification of complex mixtures separated by two-dimensional chromatography. However, for chemometric resolution and quantification of four-way data Balley and Rutan (2011) showed the need for more sophisticated algorithms (section 2.3.3.3)

2.3.2. Inverted watershed algorithm

A completely different approach for peak detection in two-dimensional chromatography is the use of software able to deal with two-dimensional data set using image analysis tools to extract the required information. To perform this task, Reichenbach *et al.* (2004) suggests the use of an inverted version of the watershed algorithm, also

known as the drain algorithm. This algorithm, which can also be used as a topographical tool, assumes that the surface of analysis has its highest point at a "mountain". When it "rains" in the "mountain", the water tends to go down throughout the cliff, and producing puddles of water surrounding the various "mountains". This algorithm is capable of foreseeing the movements of this rain and consequently delineates the peaks from an image (Reichenbach *et al.*, 2004). Conceptually, the algorithm finds the highest peak and after that it will identify the neighbouring pixels until reaching the background (Latha *et al.*, 2011). After the detection is complete, each two-dimensional peak identified is, in fact, a group of pixels. Unfortunately, the inverted watershed algorithm is very sensitive to noise and artifacts, which can cause a multi-peak detection in a single peak situation. However the application of smoothing tools to the image of the two-dimensional chromatogram may reduce the effects of these artifacts.

Once the peak detection is accomplished, then the statistical properties of the peak, such as the number of pixels (area of the peak) and the sum of pixels (volume of the peak), can be used to measure the characteristics of the two-dimensional peaks, such as symmetry, orientation or eccentricity and also to obtain quantitative information. Finally, Reichenbach *et al.* (2004, 2009) suggest the use of an algorithm for chemical identification by template matching. This algorithm uses the statistical properties of the peak to compare the sample peaks with statistical properties of several peak templates in order to find a match. While this algorithm can be used for peak detection and integration in two-dimensional chromatography, a study carried out by Vivó-Truyols and Janssen (2010) has shown that the probability of failure of this algorithm under normal conditions in GC × GC is around 15–20% when compared with the two-steps algorithm suggested by Peters *et al.* (2007a). Such a high percentage value could be a discouraging factor for a wider use of such method in Analytical Chemistry. However, Latha *et al.* (2011) tested 1000 cases for each set of parameter values, in order to compare the detection performance in both algorithms after a skew correction for various parameters, such as, different levels of noise, peak widths, and retention-time. The experiments allowed concluding that after a skew correction, the inverted watershed algorithm showed better results than the two-step algorithm. Furthermore, the accuracy of both algorithms decreases as the peak width and the noise increase, even after shift correction. Therefore, there is a need for improving the noise suppression techniques in order to achieve better results in both detection algorithms.

2.3.3. Multi-way chemometric methodologies

The increasing hyphenation of multi-array detectors such as DAD and MS detector to chromatographic systems, has brought the need for developing and/or adapting more sophisticated techniques, even for the one-dimensional chromatographic systems, in order to identify and resolve the overlapped peaks in the resulting multidimensional data structures (Amigo *et al.*, 2010). One of the legacies of one-dimensional chromatography to deal with those types of data structures is the adoption of chemometric methods for the discrimination of samples, depending on the patterns of both the chromatographic retention times and the spectral characteristics (Porter *et al.*, 2006). There is a general trend to use more and more complex chemometric algorithms to extract information from the experimental data sets, wich can attain very high dimensions and contain huge amount of data points, namely when using multichannel detectors and sample replicates. Many of these methods and their variants have already been successfully applied to two-dimensional chromatography (both LC × LC and GC × GC); as such are the cases of the Generalized Rank Annihilation Method (GRAM), Parallel Factor Analysis – Alternating Least Squares (PARAFAC-ALS), and Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS).

This family of chemometric deconvolution techniques are already widely used to deal with overlapped signals from data acquired by two-dimensional chromatography coupled to various detectors in quantitative analysis of complex samples. Their application may not be straightforward, because they are based on advanced mathematical concepts and therefore some caution should be exerted before direct application such tools from the chemometric field.

2.3.3.1. Parallel factor analysis model

The PARAllel FACtor analysis (PARAFAC) model has its origins in psychometrics sciences, but it has long been exploited in chemometrics and related areas to

deal with the increased complexity of the multi-way data sets. This is an iterative and powerful method that has already been proven useful in deconvolution and quantification in two-dimensional-chromatography (Bruckner *et al.*, 1998, 2000a, Fraga *et al.*, 2000b, 2001, 2005, Hoggard and Synovec, 2007, 2008, Latha *et al.*, 2011, Porter *et al.*, 2006, 1999a, Prazen *et al.*, 1999b, Sinha *et al.*, 2003, 2004a, 2004b, van Mispelaar *et al.*, 2003, Xie *et al.*, 2003). The PARAFAC model applied to a three-way data array can be described as follows:

$$\mathbf{R} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \mathbf{E} \qquad \textbf{(2)}$$

where **R,** in two-dimensional chromatography, is the instrumental response matrix, $\mathbf{x}_n$ is the second dimension chromatographic profile of each factor (N), $\mathbf{y}_n$ is the first dimension chromatographic profile of each factor (N), $\mathbf{z}_n$ is the detector response for each factor (N), and **E** is the error matrix with same size of **R** matrix (Hoggard and Synovec, 2007). Fig. II-8 shown a schematic representation of a PARAFAC model for 3 factors (N=3).

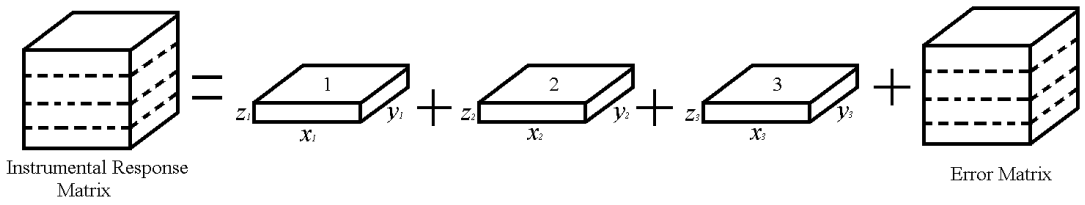

Fig. II-8. Schematic representation of a PARAFAC model with 3 components peak (visualization inspired by Bro (1997)).

Several algorithms have been described for finding the parameters $\mathbf{x}_n$, $\mathbf{y}_n$, and $\mathbf{z}_n$ and consequently fitting the PARAFAC model. These algorithms can be classified in three groups (Tomasi and Bro, 2006): a) non-iterative eigenvalue-based methods, such as Generalised Rank Annihilation method (GRAM) and the Direct Trilinear Decomposition method (DTLD); b) alternating algorithms, such as the Alternating Least Squares (PARAFAC-ALS) and the Self Weighted Alternating Trilinear Decomposition (SWATLD); and c) derivative-based methods. The GRAM and PARAFAC-ALS algorithms have been the most extensively discussed and applied to two-dimensional chromatography.

2.3.3.1.1. Generalized rank annihilation method

The GRAM has been developed by Sanchez and Kowalski (1986) based on other rank annihilation methods in order to deal with the analytical problem of detecting and quantifying one or a few components of interest in a complex mixture without the need for resolving the rest of the sample components. In practical terms, GRAM is a non-iterative eigenvalue-based method which needs two chromatograms: the sample two-dimensional chromatogram and a two-dimensional chromatogram obtained from one or more components of interest at well-defined concentrations. After that GRAM compares both two-dimensional matrices in order to provide the pure elution profiles as well the relative concentration of the analyte in the sample (Fraga *et al.*, 2000b, Xie *et al.*, 2003).

The comparisons of chromatograms performed by GRAM requires that both two-dimensional matrix must be stacked to generate a three-way data (Fraga and Corley, 2005). Consequently, it is necessary to ensure that both matrices have the same size in each row of the sample matrix corresponding to the second dimension in the calibration matrix, and also for each column of the sample matrix corresponding to the first dimension in the calibration matrix (Bruckner *et al.*, 1998). For this reason, this method only supports the analysis of bilinear samples signals (signal represented by the product of two vectors) (Fraga and Corley, 2005) and consequently does not allow the analysis of data structures of higher dimensions.

The GRAM was the first deconvolution method used in two-dimensional separations and in spite of being widely used, both in GC × GC and LC × LC (Bruckner *et al.*, 1998, Fraga *et al.*, 2000a, 2000b, 2001, 2005, Prazen *et al.*, 1999a, 1999b, Sinha *et al.*, 2003, Xie *et al.*, 2003), there are some issues that must be taken into account for ensuring a correct application. Firstly, it is necessary to make sure that the detector gives always a linear response to the variation of concentration in order to allow a proper GRAM comparison of results. Secondly, it is necessary that the peaks associated with the components of interest have the same retention time and the same peak profile in both two-dimensional chromatograms (sample and standard) (Bruckner *et al.*, 1998). Although the non-interactivity of the method justifies the fastness of the calculation performed for

reaching the results calculation, they tend to be worse than the results given by more interactive methods such as the PARAFAC-ALS.

2.3.3.1.2. Parallel factor analysis alternating least squares

The alternating least squares (ALS) was the first algorithm used to fit the PARAFAC model, since it is able to handle unresolved chemical components in three-way or higher-order data arrays (Bro, 1997). PARAFAC-ALS, which has been already used in GC × GC and LC × LC (Fraga and Corley, 2005, Hoggard and Synovec, 2007, 2008, Porter *et al.*, 2006, Sinha *et al.*, 2004a, 2004b, van Mispelaar *et al.*, 2003), is not a completely automatic method such as GRAM. Before fitting the model, it requires inputs from the user in terms of initial factor estimation and the definition of some constraints. However, selecting the appropriate number of factors in the PARAFAC model can be a real hard task to achieve. Usually, this number is given by the sum of interferences and the analytes present in the two-dimensional chromatogram, however, it is very difficult to forecast *a priori* this number for complex mixtures, particularly in the presence of either a low signal-to-noise ratio or overlapping peaks (Hoggard and Synovec, 2007). In order to deal with this limitation, Hoggard and Synovec (2007) suggested an algorithm capable of automatically select the number of factors to be used in a PARAFAC model applied to comprehensive two-dimensional separations using multichannel spectral detection: in this case a GC × GC-TOFMS.

Furthermore, there are non-negative constraints since all chromatographic signals are positive in relation to the mobile phase, and also the unimodality because chromatographic peak shapes are expected to be unimodal, that is, they usually are Gaussian-like with some fronting and/or tailing.

The ALS algorithm attempts, in each iteration, to improve the fitting of the PARAFAC model until reaching the global minimum which is the least-squares solution. This iteration process is one of the best qualities of the algorithm, but leads to one of its greatest drawbacks: the time spent on number crunching, especially with high number of variables (Bro, 1997). One of the greatest advantages of PARAFAC-ALS method over the

GRAM method is its capacity to resolve and quantify the chemical components of interest, taking into account only the sample information without the need for a standard chromatogram or multiple replicates (Porter *et al.*, 2006). The application of the PARAFAC-ALS method for resolution and quantification of targeted analytes from a four-way data array obtained by LC × LC-DAD has already been reported by Porter *et al.* (2006). Finally, the quantification can be accomplished by summing the outer product of the different dimensions of the chosen factor in the corresponding PARAFAC model across all points, thus yielding a scalar. This has already been shown by Hoggard and Synovec (2007) in a PARAFAC of target analytes in GC × GC-TOFMS data.

2.3.3.2. Target finder algorithms for PARAFAC-ALS

Since in most cases, the algorithms based on the PARAFAC model consume a significant amount of computing resources, there have been several attempts to find algorithms capable of finding targeted components in the two-dimensional chromatographic data set. These algorithms have, as their main feature, a process of screening a two-dimensional peak to find the compound of interest in a short period of time. After the compound of interest has been found, then the chromatographic sub-region of this compound can be subjected to an algorithm, such as the PARAFAC-ALS, in order to deconvolute the pure component signals to confirm the result and produce quantitative information.

2.3.3.2.1. Window target testing factor analysis

The proliferation of two-dimensional chromatographic techniques coupled to multichannel detectors has led to the development of techniques of data analysis and algorithms robust enough to deal with very large multi-dimensional data arrays. Porter *et al.* (2006) suggested the application of an algorithm capable of a fast qualitative screening of metabolites in LC × LC-DAD data sets using the window target testing factor analysis

(WTTFA) to confirm the presence or absence of an analyte. The WTTFA algorithm used for data analysis was forwarded by Lohnes *et al.* (1999), and it is based on a search for a region in the sample's chromatogram that closely resembles a known standard spectra of compounds, producing then the retention time if a match is found. The WTTFA algorithm starts to perform a singular value decomposition in a small retention time window of a one-dimensional chromatogram. The resultant spectral matrix is truncated according to the user input value of the maximum number of components estimated in the window. After that, spectra of the standard compounds are projected into the subspace described by the analysed spectra and the correlation is calculated. Finally, after the previous steps have been accomplished, the window is incremented by one time unit, and the algorithm repeats this procedure for all windows until all possibilities are tested (Porter *et al.*, 2006).

2.3.3.2.2. DotMap algorithm

The DotMap algorithm has been developed by Sinha *et al.* (2004c) for identifying spectra in GC × GC-TOFMS complex data matrices that are similar to the target spectra of interest. Firstly, there is a pre-treatment of the spectra data of the compound under study and also of the two-dimensional chromatographic data: scaling, weighting and normalization of the data, besides correction of the baseline. The spectral information of the compound under study can be previously acquired from standards or from available libraries of spectra, such as the NIST02. Afterwards, the algorithm computes the dot product "·" of the mass spectrum of the compound of interest with each mass spectrum point from the complete or partial two-dimensional chromatogram, as follows:

$$\left(\frac{m\sqrt{A_d}}{\sum m\sqrt{A_d}}\right)\cdot\left(\frac{m\sqrt{A_u}}{\sum m\sqrt{A_u}}\right) \quad \textbf{(3)}$$

where A_d is the abundance of m/z signals at each point in the two-dimensional chromatogram, A_u is the abundance of m/z signals of the compound of interest, and m is the vector containing m/z values used for weighting the signals.

The dot product result is then compared with a threshold value defined by 90% of the maximum dot product above the median of all dot products in the raw data, and a contour plot is generated with the location of the results above this threshold and the maximum value is extracted (Sinha *et al.*, 2004c). Finally, the data extracted is checked against a traditional mass spectra library in order to verify that the analysis has been well performed. This same algorithm has been evaluated by Hope *et al.* (2005) for locating analytes of interest based on mass spectral similarity in data collected using GC × GC-TOFMS.

2.3.3.3. Multivariate curve resolution with alternating least squares

Bailey and Rutan (2011) have recently developed another methodology to deal with data from two-dimensional chromatography associated to multichannel detection applied to complex mixtures. In this work, urine samples were analysed in a LC × LC-DAD system with replicates, producing a four-way data set. The aim of the work was to resolve and quantify the non-targeted overlapped compounds. In order to extract the maximum information from the complex data samples, Bailey and Rutan (2011) developed a method that combines an interactive key set factor analysis (IKSFA) technique with an "in-house" MCR-ALS algorithm with spectral selective constrains. Firstly, the two-dimensional chromatograms are divided in sections due to the complexity of the whole sample and also due to regions of detector saturation. In the section to be analysed, the IKSFA starts by the determining the number of unique spectra and estimates the spectral initial guess for the next step. After that, the MCR-ALS is applied in order to provide a resolution of the spectral different components, using non-negativity and spectral selectivity constraints. Finally, the relative concentrations are found using manual baseline integration and the percentage of relative standard deviation (*%RSD*) values are determined by comparison with standard mixtures and control samples.

Bailey and Rutan (2011) also highlighted that this algorithm does not assume multilinearity of data, which means that it can be applied when significant retention time deviations occur between samples, in both chromatographic dimensions. The algorithm is

insensitive to shifts of retention time and distortions of peak shape, and consequently does not require pre-alignment of the data before application. The use of the non-negativity and selectivity constraints are enough to obtain results fit for purpose. According to the authors (Bailey and Rutan, 2011) this lack of multilinearity even makes this algorithm more precise than the PARAFAC-ALS algorithm used by Porter *et al.* (2006). Finally, the main drawback of this method is its lack of full automation, since it requires some user intervention (Allen and Rutan, 2011, Bailey and Rutan, 2011). However, according to the authors (Bailey and Rutan, 2011), this intervention is easy to perform and very fast to accomplish.

2.4. From one-dimensional to two-dimensional chromatography: an extension of the concept of resolution

Apart from the concepts already discussed for the pre-treatment, detection and quantification of peaks, there are some other concepts from the one-dimensional chromatography that could be extended to multidimensional chromatography, namely for the case of two-dimensional chromatography, such as the measurement of peak overlapping and resolution. The extension of these concepts to two-dimensional chromatography implies the development of other concepts, such as the measurement of two-dimensional retention time, and the concept of peak vicinity.

2.4.1.1. Retention time in two-dimensional chromatography

The major obvious difference between one- and two-dimensional chromatography is the existence of a second dimension, and the first difficulty to overcome in two-dimensional chromatography is how to deal simultaneously with the two dimensions. One way to solve this problem is to work with Euclidean distances. This metric, based in the Pythagoras theorem, finds the distance between two points considering that this distance is

the hypotenuse of a right triangle whose sides are the *X* and *Y* coordinates. In practical terms, as shown in Fig. II-9, the distance $\mathbf{d_{0\text{-}1}}$ between the origin and the maximum of peak **1** is given by:

$$d_{0-1} = \sqrt{(X_1)^2 + (Y_1)^2} \quad \textbf{(4)}$$

where X_1 is the retention time of the peak **1** in the first dimension and Y_1 is retention time of the peak 1 in the second dimension. With the use of Euclidean distance it is possible to replace the two chromatographic retention times of the two-dimensional peak by one single metric, as in the case of one-dimensional chromatography.

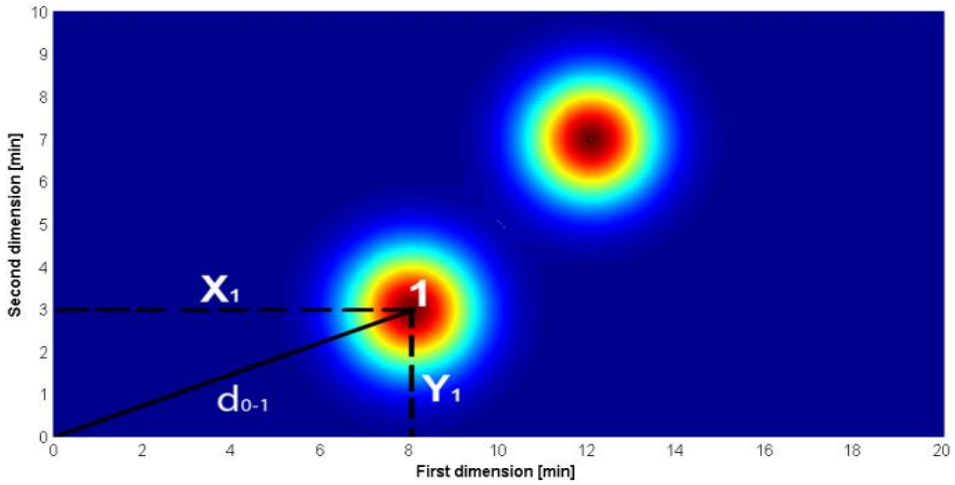

Fig. II-9. Determination of Euclidean distance for peak **1** in a two-dimensional chromatographic map.

2.4.1.2. The concept of peak vicinity

In one-dimensional chromatography, the distribution of peaks occurs only along one time axis and, consequently, it allows the existence of only two neighbouring peaks for each target peak: the peaks eluted just before and immediately after. In this sense, there are only two peaks that can have any degree of overlapping with the target peak. On the other hand, in two-dimensional chromatography, the peaks are spread all over a surface defined by the two time axis, and there may be several neighbouring peaks surrounding a target peak (Guiochon *et al.*, 2008, Peters *et al.*, 2007b).

For the purpose of calculating the resolution between peaks, Peters *et al.* (2007b) suggested that the only concern should be the study of the interaction between two consecutives peaks, thus concluding that resolution is meaningful only if no other interfering peak is in between the two peaks of interest. Therefore, prior to the resolution measurements, it becomes necessary to find the "peak vicinity", and for that it is necessary to define the two-dimensional peak regions for all the peaks of the chromatogram. This region is composed of all regions of the one-dimensional peaks, as defined by the peak detection algorithm of Peters *et al.* (2007a), and used for merging the one-dimensional into two-dimensional peaks. In order to advance in the definition of "peak vicinity", that is, to establish which peaks share the same neighbouring effect, Peters *et al.* (2007b) suggested that two peaks are neighbours if after plotting the non-interpolated trajectory profile line between two sets of two-dimensional peak clusters, there are no other peak region crossing that trajectory. As shown in Fig. II-10A, and since the profile line does not cross any other peak region, the vicinity of the two peaks is validated and the resolution can be measured. On the other hand, as shown in Fig. II-10B, if another peak region (peak **2**) has been found crossing the trajectory line between the two peaks (peak **1** and peak **3**) then the vicinity concept between these two peaks does not apply.

As highlighted by Peters *et al.* (2007b), a third case of interference can be found when between two peak clusters there is some profile trajectory line that crosses a peak region although the others do not. Fig. II-10C is an example of this third case, where there are two profile lines (represented in white) between peaks **1** and **3** that cross the region of peak **2,** although the other trajectory lines (represented in black) do not. In this third case, peaks **1** and **2** are considered as neighbours, and peak **2**, in practical terms, is not considered as an interfering peak.

The number of interfering peaks in two-dimensional chromatography of complex mixtures can be extremely large, and Peters *et al.* (2007b) suggested the use of a threshold value below which such interfering peaks may be or should be neglected. Therefore, only the peaks of intensity higher than this threshold set above the background are considered for purpose of quantification, which represents a limitation for low-concentration compounds.

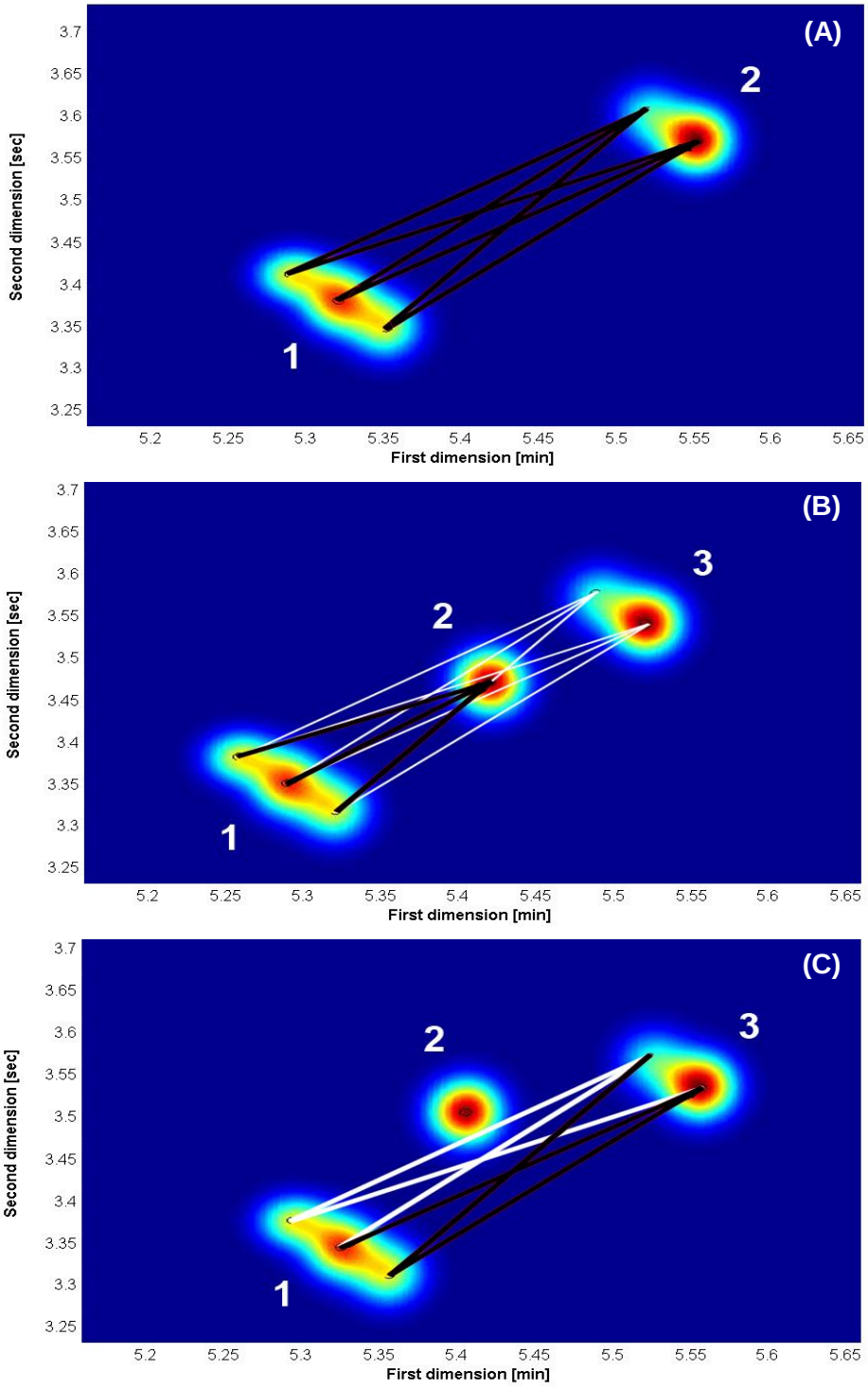

Fig. II-10. Representation of peak vicinity in three scenarios: (**A**) without an interfering peak; (**B**) with an interfering peak; and, (**C**) with a partial interference of another peak (visualization inspired by Peters *et al.* (2007b)).

2.4.2. Resolution in two-dimensional chromatography

2.4.2.1. The saddle point as a measure of overlapping

When two peaks in two-dimensional chromatography are partially overlapped, the short trajectory between these two peaks shows the characteristics of a saddle point, i. e.: it has a minimal point in one direction that is simultaneously a maximal point in another direction. As shown in Fig. II-11A, if both peaks have a Gaussian shape, then the saddle point is the minimal point located in the shortest trajectory line between the two maximal points of these peaks. This point is considered a valley point in one-dimensional chromatography (marked as a black dot). On the other hand, when the peaks do not show a Gaussian shape, it is necessary to test all possible profile trajectories between the two-dimensional peak clusters, computing all the minimal points in these lines and verifying which of these minimal points have the highest value, becoming then the saddle point.

Fig. II-11B shows the difference between considering the profile line between the two peak maxima (white line) and the profile line where the saddle point occurs (black line). The white dot in Fig. II-11B represents the minimal value occurring at the profile line between the two peak maxima, while the black dot represents the value for the saddle point. That difference as described firstly by Peters *et al.* (2007b) can be quite significant in terms of resolution.

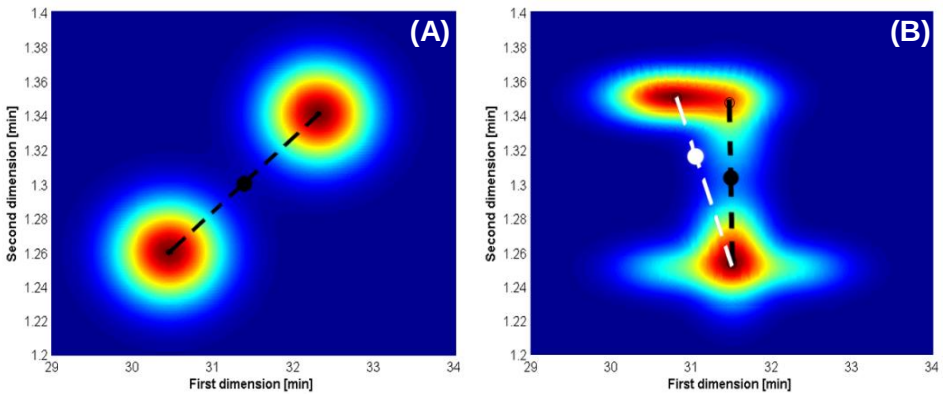

Fig. II-11. Determination of the saddle point in Gaussian peaks (**A**) and differentiation between saddle point and minimal point in non-Gaussian peaks (**B**) (visualization inspired by Peters *et al.* (2007b)).

2.4.2.2. The valley-to-peak ratio in two-dimensional chromatography

Once accomplished the determination of saddle points, and also found all the neighbouring peaks, it becomes possible to perform the calculations of resolution between the peaks (Schure, 1997). This task is based on the concept of valley-to-peak ratio, *V*, between two two-dimensional peaks, using the Kaiser's definition (Carle, 1972):

$$V = \frac{f}{g} \quad \textbf{(5)}$$

where ***g*** is distance from the baseline to the line linking the apexes of the two neighbouring peaks, and ***f*** is the distance between the height of the saddle point and this same line joining both peaks maximum, as shown in Fig. II-12.

In an ideal chromatogram, with perfect Gaussian peaks, the saddle point is exactly in a middle distance between the peaks, thus becoming easy to define the ***g*** value as the average of the two peaks height (Schure, 1997):

$$g = \frac{\left(H_{\max_1} + H_{\max_2}\right)}{2} \quad \textbf{(6)}$$

On the other hand, as shown in Fig. II-12, the following relationship can be defined:

$$f = g - H_v \quad \textbf{(7)}$$

where, H_v is the height of the saddle point (Schure, 1997).

However, in practical cases, perfect Gaussian peaks are not so common, and the saddle point is not positioned exactly at the middle distance between the peaks. Peters *et al.* (2007b) has overcome this problem by using only experimental values and geometric concepts to define the values of ***g.***

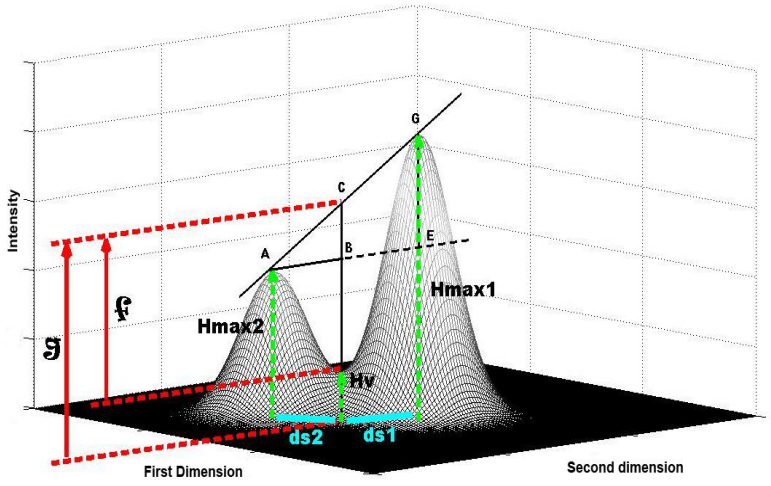

Fig. II-12. Schematic diagram of an aid for calculation of the valley-to-peak ratio between two overlapping peaks in two-dimensional chromatography.

According to Fig. II-12 it is possible to define two triangles: *ABC* and *AGE*. In these two triangles it is possible to calculate the side values using experimental chromatographic results as follows:

$$\overline{AB} = d_{s_2} \qquad \overline{AE} = d_{s1} + d_{s_2} \qquad \overline{GE} = h_{\mathbf{max}_1} + h_{\mathbf{max}_2}$$

On the other hand, both triangles share the same acute angle: therefore, the ratio between the opposite leg and the adjacent leg are equal and, in terms of sin θ, the following relationships are possible to be computed:

$$\frac{\overline{AB}}{\overline{CB}} = \frac{\overline{AE}}{\overline{GE}} = sin\theta \qquad \textbf{(8)}$$

Consequently, it is possible to define the remaining vector $\overline{CB}$ by the following equations:

$$\overline{CB} = \frac{\overline{GE}}{\overline{AE}} \times \overline{AB} \qquad \textbf{(9)}$$

$$\overline{CB} = \frac{h_{\mathbf{max}_1} - h_{\mathbf{max}_2}}{d_{s_1} + d_{s_2}} \times d_{s_2} \qquad \textbf{(10)}$$

Then the value of g can be calculated as:

$$g = \overline{CB} + h_{\mathbf{max}_2} \qquad \textbf{(11)}$$

Substituting the terms already known, Equation (11) takes the following form:

$$g = \frac{\left(h_{\mathbf{max}_1} - h_{\mathbf{max}_2}\right) \times d_{s_2}}{d_{s_1} + d_{s_2}} + h_{\mathbf{max}_2} \quad \textbf{(12)}$$

$$g = \frac{\left(h_{\mathbf{max}_1} - h_{\mathbf{max}_2}\right) \times d_{s_2} + h_{\mathbf{max}_2} \times \left(d_{s_1} + d_{s_2}\right)}{d_{s_1} + d_{s_2}} \quad \textbf{(13)}$$

$$g = \frac{h_{\mathbf{max}_1} \times d_{s_2} - \cancel{h_{\mathbf{max}_2} \times d_{s_2}} + h_{\mathbf{max}_2} \times d_{s_1} + \cancel{h_{\mathbf{max}_2} \times d_{s_2}}}{d_{s_1} + d_{s_2}} \quad \textbf{(14)}$$

Finally, it is possible to compute the value of g using only experimental results by means of the following equation (Peters *et al.*, 2007b):

$$g = \frac{h_{\mathbf{max}_1} \times d_{s_2} + h_{\mathbf{max}_2} \times d_{s_1}}{d_{s_1} + d_{s_2}} \quad \textbf{(15)}$$

where d_{s1} and d_{s2} are the distances between the retention time of each peak and, that of the saddle point, while H_{max1} and H_{max2} represent their respective heights. This procedure for the calculation of the valley-to-peak ratio was firstly suggested by Peters *et al.* (2007b) and it allows an easy way to measure the overlap of two peaks of any shape, using only the experimental raw data from the two-dimensional chromatogram.

2.4.2.3. Measuring the resolution

The last step in the estimation of resolution it is obviously the development of an equation for describing the separation quality between peaks.

Schure (1997), based on the work of Giddings (1990), proposed a metric for measuring the resolution (*Rs*) of Gaussian peaks in two-dimensional chromatography. Schure (1997) starts by considering the equation for describing two one-dimensional Gaussian peaks (here designated as peak 1 and 2):

$$h(t) = h_{\max_1} \times \exp\left[\frac{-\left(t-\bar{t}_1\right)^2}{2\sigma_1^2}\right] + h_{\max_2} \times \exp\left[\frac{-\left(t-\bar{t}_2\right)^2}{2\sigma_2^2}\right] \quad \textbf{(16)}$$

where h_1 and h_2 are the peak heights, t_1 and t_2 are retention times of the peaks, and σ_1 and σ_2 are the standard deviations.

Subsequently, and because the peaks are Gaussian, the retention time of the valley point, which is the average of the retention times of both peaks $\left(\bar{t}_v = \left(\bar{t}_1 + \bar{t}_2\right)/2\right)$, is replaced in the previous Equation (16) to determine the heights on this point :

$$h(\bar{t}_v) = h_{\max_1} \times \exp\left[-\frac{1}{2} \times \frac{-\left(t-\bar{t}_1\right)^2}{2\sigma}\right] + h_{\max_2} \times \exp\left[-\frac{1}{2} \times \frac{-\left(t-\bar{t}_2\right)^2}{2\sigma}\right] \quad \textbf{(17)}$$

This equation can be reduced to the following final expression:

$$h(\bar{t}_v) = \left(h_{\max_1} + h_{\max 2}\right) \times \exp\left[-\frac{1}{2} \times \left(\frac{\bar{t}_2 - \bar{t}_1}{2\sigma}\right)^2\right] \quad \textbf{(18)}$$

The definition of valley-to-peak ratio used for a Gaussian peak (Equation 6 and 7) can also be applied for the estimation of resolution using only experimental values:

$$V = \frac{g - h_v}{g} = \frac{\frac{h_{\max_1} + h_{\max 2}}{2} - \left(h_{\max_1} + h_{\max 2}\right) \times \exp\left[-\frac{1}{2} \times \left(\frac{\bar{t}_2 - \bar{t}_1}{2\sigma}\right)^2\right]}{\frac{h_{\max_1} + h_{\max 2}}{2}} \quad \textbf{(19)}$$

$$V = 1 - 2\exp\left[-\frac{1}{2} \times \left(\frac{\Delta\bar{t}}{2\sigma}\right)^2\right] \quad \textbf{(20)}$$

Using the traditional one-dimensional resolution equation formula, $Rs = \Delta\bar{t}/4\sigma$, the valley-to-peak ratio can be transformed into the following equation:

$$V = 1 - 2\exp(-2Rs^2) \quad \textbf{(21)}$$

Finally, and by manipulating Equation (21), it is possible to compute the resolution using the value of the valley-to-peak ratio (Schure, 1997):

$$Rs = \sqrt{-\frac{1}{2}\ln\left(\frac{1-V}{2}\right)} \qquad \textbf{(22)}$$

Equation (22) has been generalized by *Peters et al.* (2007b) for the determination of the valley-to-peak ratio between non-Gaussian two-dimensional peaks, and it allows the description of the overall separation regardless of the peak shape. Such a generalization constitutes a huge advance for evaluating the chromatographic performance in the two-dimensional chromatography.

Schure (1997) also presented a compilation of resolution metrics proposed by other authors. The first formula was compiled from the work of Davis (1991) and Shi and Davis (1993):

$$Rs(\theta) = \frac{\delta_t \sqrt{\gamma^2 \sin^2\theta + \cos^2\theta}}{4\sigma_x} \qquad \textbf{(23)}$$

where δ_t are the deviations between peaks using the Euclidean distance, σ_x are the standard deviations of the x peak zone, γ is the ratio between the x and y standard deviations, and θ is the angle between the line that links both peaks and a parallel line to the x axis, as schematically represented in Fig. II-13.

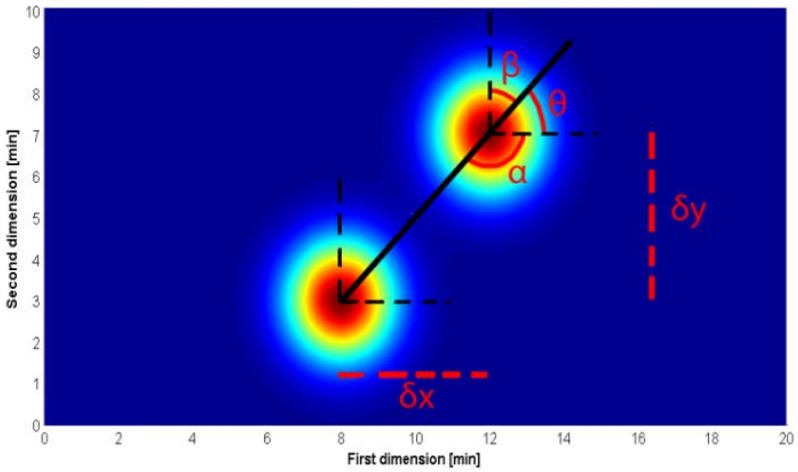

Fig. II-13. Representation of a two-dimensional chromatogram zone for the estimation of resolution using Equation (23).

Another resolution metrics presented by Schure (1997) was proposed by Giddings (1990), and it considers than the total resolution follows the Euclidean norm:

$$Rs = \sqrt{Rs_x^2 + Rs_y^2} = \sqrt{\left(\frac{\delta_x^2}{16\sigma_x^2}\right) + \left(\frac{\delta_y^2}{16\sigma_y^2}\right)} \quad \textbf{(24)}$$

where σ_x and σ_y are the standard deviations of peaks *x* and *y*, respectively; and δ_x and δ_y are calculated as $\left(\bar{t}_{2,x} - \bar{t}_{1,x}\right)$ and $\left(\bar{t}_{2,y} - \bar{t}_{1,y}\right)$, respectively.

2.5. Data analysis software

The huge amount of data sets produced by two-dimensional separations make their analysis almost impossible without using some type of computer software in order to transform the two-dimensional chromatographic data into usable information. Pierce *et al.* (2008), in a recent review about advancements in comprehensive two-dimensional separations with chemometrics, concluded that most of the available commercial and public domain software has been adapted from one-dimensional chromatography. One of the few examples of commercial software developed and available for GC × GC is the software system developed as a spin off at the University of Nebraska-Lincoln (Reichenbach *et al.*, 2003, 2004, 2005). The *GC Image* incorporates the inverted watershed algorithm and allows a digital image processing for visualization, processing, analysis and reporting of GC × GC chromatographic data. Recently, a version of this software has been developed and released for dealing with LC × LC data (Reichenbach *et al.*, 2009). Quite often, many researchers have developed their own algorithms, usually in *MATLAB* (Mathworks, Natick, MA, USA), a well-known commercial software for numerical computation. Several of this algorithms and toolbox available in *MATLAB* language for analysis of multi-way data speeded all over the internet. An example of this source code is the N-way toolbox available at *http://www.models.kvl.dk/source/* (Andersson and Bro, 2000) . There are no reports comparing the performance of different commercial software available but there are a few studies comparing different algorithms already incorporated in commercial software packages (Latha *et al.*, 2011, Vivó-Truyols and Janssen, 2010).

2.6. Conclusions

Data processing of comprehensive two-dimensional chromatography is a rapid evolving subject since there is a general lack of commercial software associated to analytical instrumentation. Although the first algorithms developed for data processing in two-dimensional chromatography were generalizations of concepts from one-dimensional chromatography, nowadays there are already methods for non-targeted and targeted analyses that deal with the two-dimensional chromatographic peaks as a whole, without the need for merging the several one-dimensional sections of a single two-dimensional peak. Furthermore, there are concepts from one-dimensional chromatography, such as peak purity and chromatographic response function that have not been attempted in two-dimensional chromatography. Such work may provide new insights and progress into data processing.

The emergence of multichannel detectors will easily lead to obtaining N-way data that need expertise drawn from chemometrics for proper data processing and attaining adequate information for analytical purposes. Researchers often choose to develop their own algorithms and consequently it becomes hard to compare the results obtained. Besides, such an expert knowledge may take some time to be embedded in user-friendly software associated to the analytical instrumentation available in chemical laboratories. The trend on this matter appears to be leading to sophisticated algorithms such as the Bailey and Rutan (2011) MCR-ALS.

III

A new chromatographic response function for assessing the separation quality in comprehensive two-dimensional liquid chromatography

3.1. Introduction

Data handling and optimal extraction of information in two-dimensional chromatography, namely the strategies related to data treatment/resolution measurement and the design of optimal two-dimensional separation conditions, still are one of the major drawbacks that have impaired a wider application of this analytical technique. In the particular case of LC × LC, the design of a suitable experimental procedure is an enormous analytical challenge that still lacks the use of an important requirement in any chromatographic optimization procedure: a mathematical function to measure and map the two-dimensional separation quality. In fact, no reference could be found in the literature that addresses the definition and use of a CRF for qualifying the resolution level associated with any two-dimensional peak arrangement. So far, the most known composite CRFs described in the literature were proposed for optimization purposes in LC. Besides a quality index for the separation, the existing CRFs in LC comprise other important secondary criteria, namely analysis time, number of detectable peaks, and/or robustness (Berridge, 1982, Bolanča and Cerjan-Stefanović, 2007, Bostyn *et al.*, 2009, Bylund *et al.*, 1997, García-Álvarez-Coque *et al.*, 2006, Gheshlaghi *et al.*, 2008, Morris *et al.*, 1996, Siouffi and Phan-Tan-Luu, 2000). Each of these CRF was designed to quantify the resulting chromatograms based on the ultimate goal of the separation, and they usually require a previous knowledge on time constrains, desired peak resolution, and acceptable analysis time, and in most cases they rely on a model built on a theoretical basis (García-Álvarez-Coque *et al.*, 2006, Nikitas and Pappa-Louisi, 2009). Such requirements indicate that none of the existing CRFs is particularly suitable for mapping the quality of the chromatographic separation profiles of unknown samples. To solve this problem, Duarte and Duarte (2010) have recently proposed an alternative approach for a CRF in LC. Equation (25) was applied to complex mixtures of organic compounds, being well suited for describing the separation of peak pairs of highly unequal area, and also for overlapped and asymmetric peaks (Duarte and Duarte, 2010). Furthermore, it does not require the prior definition of an optimum and/or minimum acceptable resolution, which limits the application of any objective function to the separation of highly complex systems.

$$CRF = \sum_{i=1}^{N-1} \theta_i + N - \left(\frac{t_{R,L} - t_0}{t_{R,L}} \right) \quad \textbf{(25)}$$

where θ_i is a normalized measurement of the peak resolution using the valley-to-peak ratio for adjacent peaks, N is the number of detected peaks, $t_{R,L}$ is the retention time of the last eluted peak, and t_0 is the elution time corresponding to the column void volume.

Introducing a new CRF for the LC × LC analysis of complex unknown mixtures is not, however, straightforward, and it requires the redefinition of the conventional LC quality criteria. Using Equation (25) as a starting point for this study, the concepts of peak resolution and analysis time in the two-dimensional domain have to be revisited. In terms of a resolution measure in two-dimensional separations, as described in section 2.4.2, the existing concept has evolved from the classical definition of valley-to-peak ratio in LC and it has been applied firstly for two-dimensional Gaussian peaks (Schure, 1997) and latter for non-Gaussian-shaped two-dimensional peaks (Peters *et al.*, 2007b). In fact, nowadays, the methodology proposed by Peters *et al.* (2007b) constitute the most systematic procedure for establishing an overall quality criterion of a particular two-dimensional separation problem exhibiting non-Gaussian two-dimensional peaks. However, it has to be mentioned that in the case of the two-steps peak detection algorithm of Peters *et al.* (2007b), the "merging" step is a time consuming process, particularly when analysing large amounts of data, and the application of this peak-detection algorithm exhibits some drawbacks, particularly when dealing with closely-eluting peaks in both separation dimensions, as acknowledged by the authors.

At this point, and in spite of its popularity for peak detection, it is obvious that the two-steps algorithm does not incorporate all the solutions for every two-dimensional peak-detection problem, specially for the analysis of complex unknown samples by LC × LC. Then again, the only alternative available is the watershed algorithm (section 2.3.2) proposed by Reichenbach *et al.* (2003, 2004, 2005, 2009), but this algorithm is very sensitive to time shifts variations (Vivó-Truyols and Janssen, 2010) and need some chemometric pre-treatment algorithm to reduce this problem (Latha *et al.*, 2011). As the application of LC × LC for the analysis of complex unknown samples is still in its infancy, to find alternative and simpler approaches for peak-detection in LC × LC is still of significant interest.

Furthermore, both the "peak vicinity" concept and the estimation of the valley-to-peak ratio, based on the saddle point concept, proposed by Peters *et al.* (2007b), present some drawbacks. For instance, the "peak vicinity" concept does not take into account how far the two-dimensional peaks are separated, which means that two peaks can be considered neighbours even when located far apart from each other. This feature may be viewed as a weakness of the proposed methodology, and the estimation of the resolution between these two peaks actually becomes redundant from a "data treatment" point of view. On the other hand, as it will be demonstrated later in this chapter, the saddle point concept may fail when applied to a two-dimensional chromatogram containing a large amount of peaks with a poor resolution degree. In such situations, Equation (22) cannot be applied for the estimation of the resolution between the two-dimensional peak pairs. However, establishing the best chromatographic conditions in such situations continues to be of especial interest, which means that finding new approaches to qualify the degree of chromatographic separation becomes fundamental, and it will constitute a major breakthrough for an effective widespread application of LC × LC in the analysis of complex samples.

Therefore, in this third chapter, a new CRF for using in LC × LC of complex mixtures, evolving from the one-dimensional model recently proposed by Duarte and Duarte (2010), Equation (25), is presented. The estimation of the quality index of separation is now oriented to each individual two-dimensional peak, instead of each two-dimensional peak pair. The total number of distinguishable two-dimensional peaks and the time needed to achieve a sufficiently resolved two-dimensional chromatogram are also decisive for the global optimization process. On this regard, a new "time-economy" function is also proposed as a secondary criterion to assess the quality of separation. The performance of the developed CRF is illustrated by simulated two-dimensional chromatograms and LC × LC analysis of a mixture of four aromatic compounds, using the acetonitrile content of the mobile phase of first-dimension as the experimental variable. The advantages and drawbacks of the proposed objective function for LC × LC are also highlighted.

3.2. Development of a two-dimensional chromatographic response function

The following three consecutive major steps were considered for the development of the two-dimensional chromatographic response function (CRF_{2D}): a) two-dimensional peak detection using the interpretation of the sum of the partial second-derivatives of the two-dimensional chromatogram; b) assessment of the quality of two-dimensional peak separation using a novel two-dimensional peak purity concept; and c) set-up of the objective function incorporating all the criteria established in the previous steps. Additional details regarding each step are discussed in the following sub-sections.

3.2.1. Two-dimensional peak detection

Regardless the method used for the qualitative evaluation of a two-dimensional chromatogram, the raw data obtained in a LC × LC experiment must first be converted into a two-dimensional matrix grid using the modulation period. For large modulation periods, it is advisable to interpolate between experimental values. The two-dimensional peaks can be then detected using an algorithm based on the search for a regional maximum value. As shown in Fig. III-1, $M_{(i,j)}$ represents the value under test of the two-dimensional matrix, at the row *i* and column *j*. The basis of the algorithm is to perform a test on every and each of the matrix values, and finding out whether all the eight surrounding values are lower than the tested value. Once detected, all regional maxima contain data on the height of peaks as well as their respective two-dimensional coordinates.

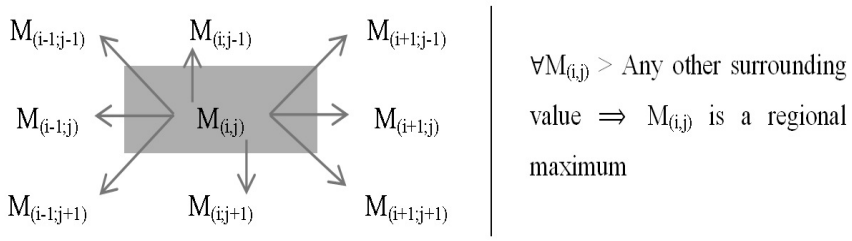

Fig. III-1. Basis for definition of the algorithm for detection of two-dimensional peaks as regional maxima.

However, when dealing with overlapped peaks there can be peaks with its maximum hidden, and consequently they cannot be detected by an algorithm of the same type as mentioned above. In such cases, peak detection can still be achieved using the interpretation of the second derivative of the chromatogram. Fig. III-2 exemplifies how the second-order derivative can be used to determine the retention time of strongly overlapping and shouldering one-dimensional chromatographic peaks. The minimum values of the negative peaks of the second derivative (Fig. III-2B) represent the retention time of the peaks appearing in the zero-order chromatogram (Fig. III-2A). In the example of Fig. III-2 there are a total of three peaks: one peak with only one identifiable maximum and two overlapping peaks with clearly visible peak maxima.

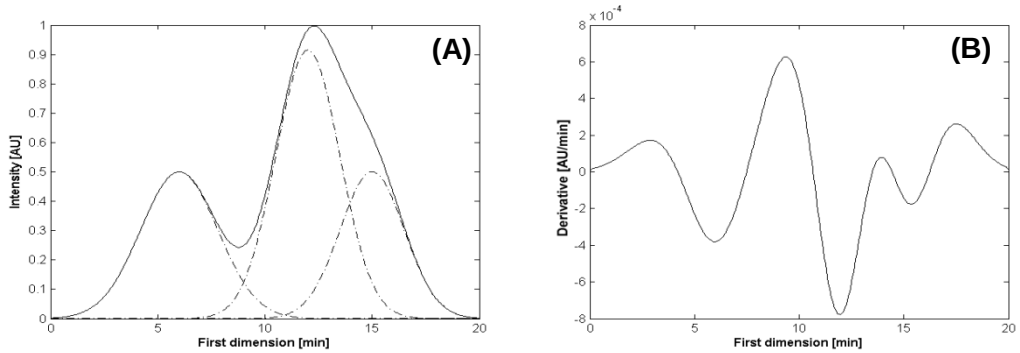

Fig. III-2. Computer simulation of overlapping one-dimensional chromatographic peaks: (**A**) composite chromatogram (solid line) and its deconvolution (dotted line) highlighting the existence of overlapped peaks; (**B**) second-order derivative of the composite chromatogram.

Computing the second-order derivative of a two-dimensional chromatogram is not, however, straightforward as in the case of a one-dimensional chromatogram. In a one-dimensional function (one variable), the derivative process takes place only along the *x* axis dimension, whereas in a two-dimensional problem (two variables) it is possible to compute partial derivatives in respect to both *x* (first dimension) and *y* (second dimension) axes. The partial second-derivative of a two-dimensional function usually generate the second-derivative in respect to *x*, $\partial z/\partial x^2$, the second-derivative in respect to *y*, $\partial z/\partial y^2$, and the second-derivative in respect to *x* and *y*, $\partial z/\partial x \partial y$. The detection of a two-dimensional peak in cases of strongly overlapping peaks can be achieved by plotting the sum of the two partial second-derivatives, $\partial z/\partial x^2$ and $\partial z/\partial y^2$. The *x* and *y* coordinates of the minimum point of this two-dimensional representation corresponds to the coordinates of the maximum point in the original two-dimensional matrix. Fig. III-3 illustrates the application of this

methodology to a computer simulated two-dimensional chromatogram, which is represented as a contour plot (Fig. III-3A) and a three-dimensional surface plot (Fig. III-3B) for a better visualization of the two-dimensional peaks. The second derivative of the two-dimensional chromatogram is shown in Fig. III-3C and Fig. III-3D, where the minimum values are represented by the dark-blue colour. The automatic search for the minima can then be implemented by applying an algorithm similar to that used to detect well defined two-dimensional peaks as regional maxima (Fig. III-1), although, in this case, it has been modified in order to detect regional minima. In the example of Fig. III-3, the algorithm for peak detection was able to identify two strongly overlapping two-dimensional peaks.

Finally, it should be mentioned that the numerical calculations for producing the second-derivatives can generate minor numerical instabilities, which are directly related to the original matrix size and grid spacing. A visual inspection of both the 3D raw data and second derivatives should provide clues on whether some degree of chromatogram smoothing or even peak modelling would be advisable.

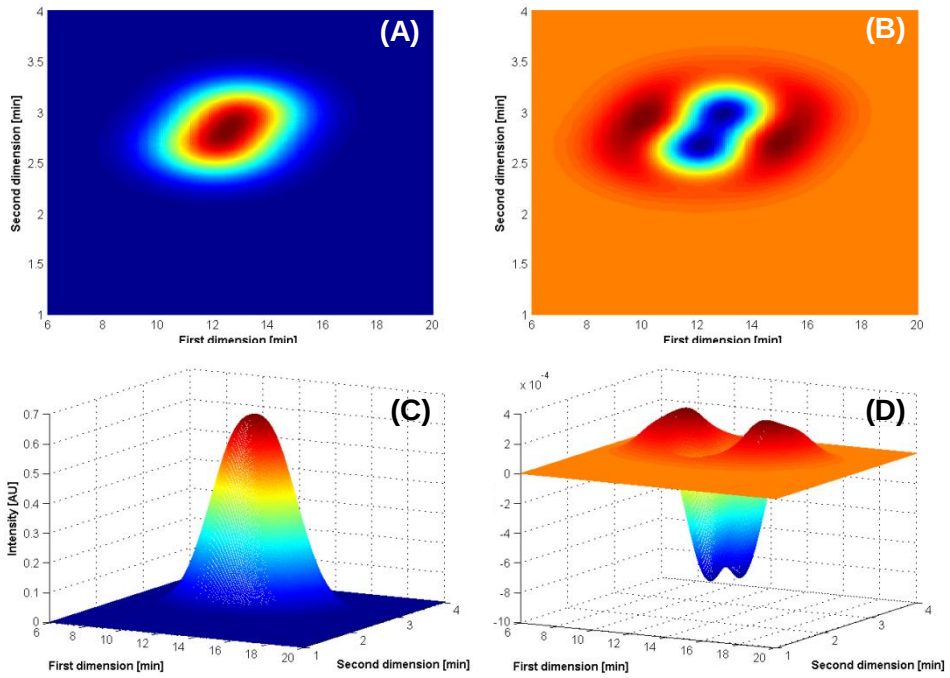

Fig. III-3. Contour plot and three-dimensional surface plot of two two-dimensional chromatographic peaks highly overlapped ((**A**) and (**B**), respectively) and the sum of their second partial derivatives ((**C**) and (**D**), respectively).

3.2.2. Assessment of the quality of separation between peaks: the concept of peak purity revisited

After determining the number of two-dimensional peaks and their respective retention times, the next step entails the evaluation of the resolution level associated to the peaks arrangement in the two-dimensional chromatogram. The resolution metrics (section 2.4.2) for comprehensive two-dimensional chromatography made available by Peters *et al.* (2007b) rely on the use of the saddle point concept. Although this concept is most suitable for expressing the degree of interaction in a system of two peaks, regardless their shape, it may fail when applied to a system containing a large amount of peaks, such as in the case of LC × LC of complex organic mixtures. The first problem emerges when attempting to apply the concept of saddle point as a measure of resolution between strongly overlapping and shouldering peaks. Fig. III-4 shows a simulation of a two-dimensional chromatogram where the proposed algorithm for peak detection identified three overlapping peaks (**A**, **B**, and **C**). Using the concept proposed by Peters *et al.* (2007b), it is possible to identify the saddle point between peaks **A** and **B** ($\mathbf{S_{AB}}$) and between peaks **B** and **C** ($\mathbf{S_{BC}}$), but no saddle point exist between the peaks **A** and **C**. This result suggest that the peak pair **A**-**C** must be excluded, meaning that the resolution between those two peaks is not calculated, which certainly hinders the assessment of the overall quality of separation of the two-dimensional chromatogram.

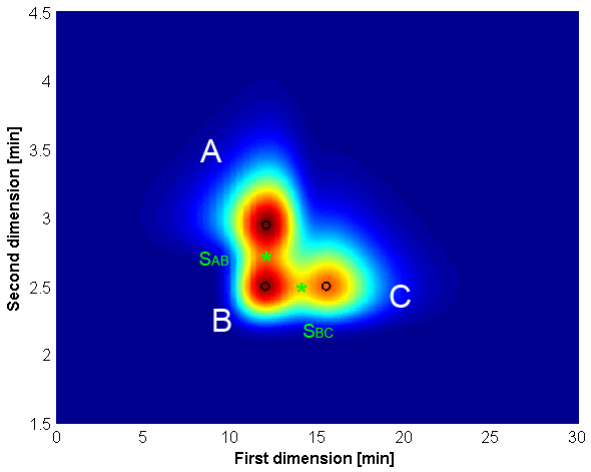

Fig. III-4. Representation of a portion of a two-dimensional simulated chromatogram, with three peaks (**A**, **B** and **C**) and two saddle points ($\mathbf{S_{AB}}$ and $\mathbf{S_{BC}}$).

The second problem that arise from the application of the algorithm proposed by Peters *et al.* (2007b) to LC × LC of complex mixtures, is related to the concept of peak vicinity. Those authors suggested that two peaks are neighbours if, after plotting the non-interpolated trajectory profiles between two sets of two-dimensional peak clusters, there is no other peak region crossing that same trajectory. However, in a region with three interconnected consecutive two-dimensional peaks, the tail of the first peak could be long enough that it may pass through the second peak and reach even the third peak. Since the vicinity is found by taking into consideration only the second peak, this two-dimensional peak arrangement implies that the interaction between the first and the third peak is not accounted for and this combination is excluded from the computation of the overall resolution of the two-dimensional chromatogram.

In order to overcome the above mentioned difficulties, an alternative approach is suggested based on the application of a peak purity concept as a measure of the quality of the LC × LC separation. The peak purity criterion has been widely used in LC and it quantifies the interference level for a given peak in the chromatogram (García-Álvarez-Coque *et al.*, 2006, Ortín *et al.*, 2011):

$$P_i = 1 - \left(\frac{A_i^{'}}{A_i} \right) \qquad \textbf{(26)}$$

where $A_i^{'}$ is the area under a given peak overlapped by the chromatogram yielded by the remaining peaks, and A_i is the total area of that peak. This criterion is a normalized measurement that ranges between 0 and 1, where the value 1 means that the peak is completely free of interference (i.e., completely resolved). In practical terms, the use of this concept in LC × LC is advantageous because its meaning is very intuitive: the obtained numerical value represents the exact portion of the interference–free two-dimensional peak (García-Álvarez-Coque *et al.*, 2006). The most important feature of this figure of merit is that the evaluation of the resolution degree of a two-dimensional chromatogram is now oriented to the estimation of the separation quality for each individual two-dimensional peak, instead of each two-dimensional peak pair as occurred in Peters *et al.* (2007b) resolution concept. This is important not only for a better implementation of a general model for the CRF_{2D}, but also to avoid the problems of the multiple vicinities between several neighbouring peaks in a two-dimensional liquid chromatogram. It should be further

mentioned that it is possible to combine all the individual values of peak purity into a single global value. In the case of full resolution of all the two-dimensional peaks, the maximum value for the sum of all individual values equals the number of two-dimensional chromatographic peaks. This maximum value is important for applications in a CRF_{2D}, since it allows identifying the highest contribution of this term to the chromatographic function and also the deviation from the theoretical and maximum achievable value.

3.2.2.1. Two-dimensional peak fitting

In one-dimensional chromatography, in case of overlapping or shouldering peaks, the original chromatographic data set do not give any information regarding the profile of each peaks, which hinders any attempt to quantify the interference level for each of those peaks and, consequently, their peak purity measure. So, the computation of the overlapped peak area and, therefore, the peak purities, requires the prediction of not only the peak location but also of its profile. Usually, this profile is computed by the simulation of each chromatographic peak from the real data using an asymmetrical peak model (Ortín *et al.*, 2011), such as the exponentially modified Gaussian, the Poisson, the Log-normal, or extreme value functions (Di Marco and Bombi, 2001). In practice, it is usually used one of the several commercial peak-fitting software packages available (Di Marco and Bombi, 2001). Fig. III-5 illustrates an example of this one-dimensional peak simulation for three overlapped peaks. It is possible to see that the straightforward methodology for calculating peak purity is to fit a model for each peak. Once a model is fitted, the degree of interference between every pair of peaks in one-dimensional chromatography can be easily calculated.

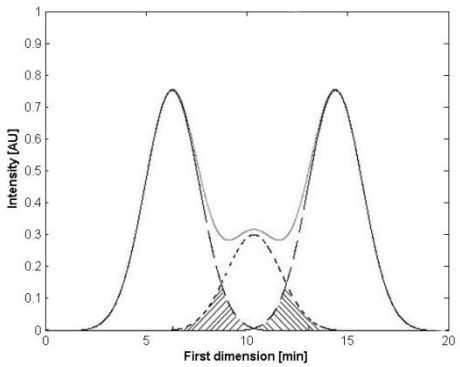

Fig. III-5. Representation of a one-dimensional chromatogram (thick line) with the representation of the one-dimensional peaks that can explain this shape (dashed line); the shadowed regions represent the degree of interference between two peaks.

In LC × LC, however, each chromatographic peak can be viewed as a surface on a three-dimensional Cartesian plane, which means that the computation of the peak purity value of a two-dimensional peak through Equation (26) requires the replacement of the areas $A_i^{'}$ and A_i by the corresponding volume values. On the other hand, dealing with raw data obtained from real experimental two-dimensional chromatographic system, in special for LC × LC, makes this fitting process very complex. In order to overcome these difficulties, and to obtain information on each individual two-dimensional peak in the chromatogram, an alternative approach is suggested based on the use of a two-dimensional mathematical model to fit the original two-dimensional chromatographic data.

For the purpose of this study, a sum of two-dimensional three parameters Log-normal functions is used to fit the overlapped peaks identified in a given section of a two-dimensional chromatogram. This function is chosen due to its relative simplicity but also its high adaptability to the fitting process. The Log-normal function has been widely used for describing chromatographic peaks in LC. There are several formulae for this function reported in the literature and all are mathematically equivalent to each other (Di Marco and Bombi, 2001). The two-dimensional Log-normal function used in this study for describing the profile of a single two-dimensional peak is defined as:

$$h_{(t_{1D},t_{2D})} = h_0 \times \exp\left\{-0.5 \times \left\{\left[\frac{\ln\left(\frac{t_{1D}}{t_{R,1D}}\right)}{\sigma_{1D}}\right]^2 + \left[\frac{\ln\left(\frac{t_{2D}}{t_{R,2D}}\right)}{\sigma_{2D}}\right]^2\right\}\right\} \quad \textbf{(27)}$$

where $h(t_{1D},t_{2D})$ is the height at time t_{1D} and t_{2D} in the first and second dimensions, respectively, h_0 is the maximal peak height, $t_{R,1D}$ and $t_{R,2D}$ is the retention time of the peak in the first and second dimensions (i.e. peak coordinates), respectively, and σ_{1D} and σ_{2D} is the standard deviation of the peak in the first and second dimensions, respectively. The fitting process will be implemented by a Trust-Region algorithm, with some pre-defined parameters, namely the peak coordinates ($t_{R,1D}$ and $t_{R,2D}$) and the range of values of the model parameters to be fitted, i.e. h_0, σ_{1D}, and σ_{2D}. The quality of the fitting can be assessed through the adjusted regression coefficient (R^2_{adj}), which represents the percentage of the variance accounted for by the fitted model. However, it should be noted that other two-dimensional mathematical models such as the Exponential Modified Gaussian and Extreme value functions, can also be used for fitting the two-dimensional peaks, and similar results as those produced by the three parameters Log-normal functions can be obtained. On this regard, it should be mentioned that this study is not meant to be a comparison of all the two-dimensional mathematical models available that can be used for this purpose, instead it focus on the development of a CRF_{2D}. Finally, once the best model is found for each of the two-dimensional peaks, then it becomes possible to calculate the two-dimensional peak purity based on the peak volume values (in the following sub-sections).

3.2.2.2. Measurement of two-dimensional peak volume for assessing the peak purity

If one considers a two-dimensional peak as a surface described by a positive function of two variables, $z = f(x,y)$, the volume of the region between the surface and the plane which contains its domain can be calculated through the definite double integral of $z = f(x,y)$. Despite of the laborious computation associated to the numerical integration of a function of two variables, this is the most simple and easiest mathematical pathway to calculate the volume under a surface on a three-dimensional plane. However, identifying the volume of an overlapping region between two or more two-dimensional peaks and calculating its numerical integral is a dreadful task since no mathematical equation for computing such volume intersection is currently available.

A possible alternative for estimating the volume of a two-dimensional peak from the obtained raw data entails the sum of all the values of the two-dimensional matrix grid and multiplying the obtained result by the spacing of the grid in both dimensions. In order to demonstrate the feasibility and accuracy of this numerical approach, a two-dimensional peak (illustrated in Fig. III-6) was generated through a two-dimensional Extreme Value type function:

$$h_{(t_{1D},t_{2D})} = h_0 \times \exp\left\{-\exp\left[-\left(\frac{t_{1D}-t_{R,1D}}{\sigma_{1D}}\right)\right]-\left[-\left(\frac{t_{1D}-t_{R,1D}}{\sigma_{1D}}\right)\right]+1\right\}\times$$
$$\exp\left\{-\exp\left[-\left(\frac{t_{2D}-t_{R,2D}}{\sigma_{2D}}\right)\right]-\left[-\left(\frac{t_{2D}-t_{R,2D}}{\sigma_{2D}}\right)\right]+1\right\} \quad \textbf{(28)}$$

where the parameters appearing in this mathematical equation have the same meaning of those described in Equation (27) and assume the following values: h_0 = 0.5 (AU), $t_{R,1D}$ = 22.0 min, $t_{R,2D}$ = 2.0 min, σ_{1D} = 1.5, σ_{2D} = 1.0, t_{1D} = 18.0 to 30.0 min with a spacing of 0.15, and t_{2D} = 1.0 to 3.0 min with a spacing of 0.05. The volume of the two-dimensional peak in Fig. III-6 was calculated by two different ways: a) using the double integral of Equation (28) over the region between the coordinates 18.0 and 30.0 in the first dimension, and coordinates 1.0 and 3.0 in the second dimension (obtained result = 0.551484); and b) using the numerical approach here suggested (obtained result = 0.551499). A relative error of 0.0025% in the prediction of the volume values by the two mathematical approaches suggests that the method here developed is highly acceptable and fit for the purpose of computing the volume of a two-dimensional chromatographic peak. The relative error can be further improved by changing the grid spacing in both dimensions, which can be done at least at two different stages of the chromatographic process: data acquisition and data treatment.

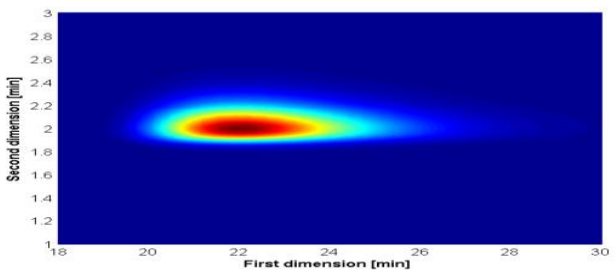

Fig. III-6. Representation of a two-dimensional peak simulated by an Extreme value peak fuction.

3.2.2.3. Measurement of the volume of a two-dimensional peak for assessing peak purity

In the case of overlapping peaks, and after calculating the total volume of the two-dimensional peaks, the following step entails the computation of the volume of the intersection region. In one-dimensional chromatography, the distribution of peaks occurs only along one time axis, and consequently it allows the existence of only a maximum of two neighbouring peaks, corresponding to the peaks eluted immediately just before and after of a given target peak. In LC × LC, however, this estimate is not an easy task because the two-dimensional peaks are spread all over the two-dimensional chromatographic surface and consequently one single peak could be interacting with more than one neighbouring peaks. To illustrate and better understand the possible cases of overlapping peaks, one can make use of Venn diagrams, as depicted in Fig. III-7, Fig. III-8, and Fig. III-9, where the circles **A**, **B**, **C** and **D** represent the two-dimensional chromatographic peaks.

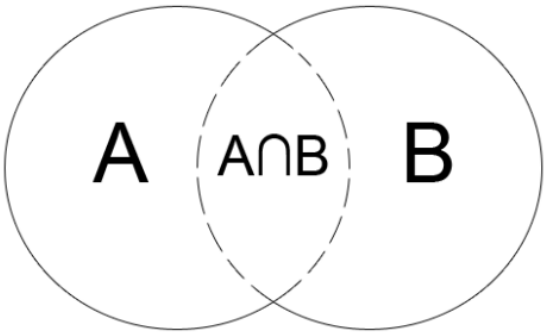

Fig. III-7. Venn diagrams representing the overlapping of two two-dimensional peaks (**A** and **B**).

In the first case, shown in the Fig. III-7 the overlapping region is the same for peaks **A** and **B**, and it reflects the volume of the intersection between these two peaks ($V_{A\cap B}$). For this case, the purity of peak **A**, P_A, and the purity of peak **B**, P_B, can be calculated through Equation (29) and Equation (30), respectively:

$$P_A = 1 - \left(\frac{V_{A\cap B}}{V_A} \right) \quad \textbf{(29)}$$

$$P_B = 1 - \left(\frac{V_{A \cap B}}{V_B} \right) \quad \textbf{(30)}$$

where V_A and V_B are the volume of peaks **A** and **B**, free from interferences, respectively. The volume of the intersection between these two peaks is calculated as follows:

$$V_{A \cap B} = V_A + V_B - V_T \quad \textbf{(31)}$$

where V_T being the total volume of the two-dimensional peaks.

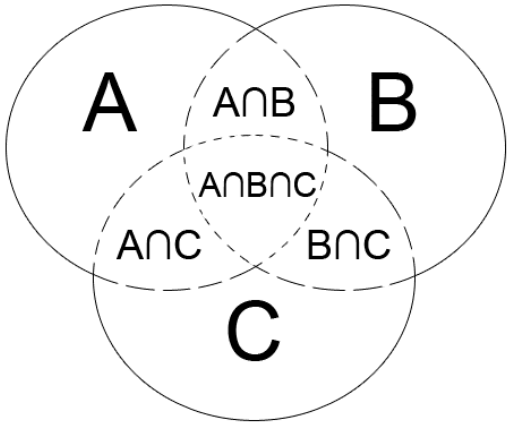

Fig. III-8. Venn diagrams representing the overlapping of three two-dimensional peaks (**A**, **B**, and **C**).

Fig. III-8 exemplifies the case when there are interferences of second order; that is, when each two-dimensional peak interacts with other two peaks, leading to the existence of an interaction region that is common to all the peaks ($V_{A \cap B \cap C}$). In this case, the value of $V_{A \cap B \cap C}$ should be removed from the computation of the overlapped volume of each peak (e.g. for peaks **A**, overlapped volume = $V_{A \cap B}$ + $V_{A \cap C}$ - $V_{A \cap B \cap C}$), and the purity of peaks **A**, **B** and **C** can be calculated through Equation (32), Equation (33) and Equation (34), respectively:

$$P_A = 1 - \left[\frac{\left(V_{A \cap B} + V_{A \cap C} - V_{A \cap B \cap C} \right)}{V_A} \right] \quad \textbf{(32)}$$

$$P_B = 1 - \left[\frac{\left(V_{A \cap B} + V_{B \cap C} - V_{A \cap B \cap C} \right)}{V_B} \right] \quad \textbf{(33)}$$

$$P_C = 1 - \left[\frac{\left(V_{A \cap C} + V_{B \cap C} - V_{A \cap B \cap C} \right)}{V_C} \right] \quad \textbf{(34)}$$

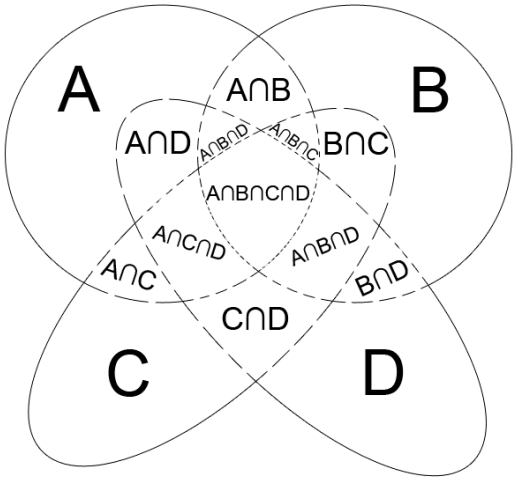

Fig. III-9. Venn diagrams representing the overlapping of four two-dimensional peaks (**A**, **B**, **C** and **D**).

Finally, Fig. III-9 illustrates the case when there are interferences of second order (**A**∩**B**∩**C**, **A**∩**B**∩**D**, **A**∩**C**∩**D**, and **B**∩**C**∩**D**), which is an overlapped region common to three of four peaks, and an interference of third order (**A**∩**B**∩**C**∩**D**) which is an overlapped region common to all the four peaks. Using the same reasoning as above for the case of three overlapping peaks, the purity of peaks **A**, **B**, **C** and **D** can be calculated as follows:

$$P_A = 1 - \left[\frac{\left(V_{A\cap B} + V_{A\cap C} + V_{A\cap D} - V_{A\cap B\cap C} - V_{A\cap B\cap D} - V_{A\cap C\cap D} + V_{A\cap B\cap C\cap D} \right)}{V_A} \right] \quad \textbf{(35)}$$

$$P_B = 1 - \left[\frac{\left(V_{A\cap B} + V_{B\cap C} + V_{B\cap D} - V_{A\cap B\cap C} - V_{A\cap B\cap D} - V_{B\cap C\cap D} + V_{A\cap B\cap C\cap D} \right)}{V_B} \right] \quad \textbf{(36)}$$

$$P_C = 1 - \left[\frac{\left(V_{A\cap C} + V_{B\cap C} + V_{C\cap D} - V_{A\cap B\cap C} - V_{A\cap C\cap D} - V_{B\cap C\cap D} + V_{A\cap B\cap C\cap D} \right)}{V_C} \right] \quad \textbf{(37)}$$

$$P_D = 1 - \left[\frac{\left(V_{A\cap D} + V_{B\cap D} + V_{C\cap D} - V_{A\cap B\cap D} - V_{A\cap C\cap D} - V_{B\cap C\cap D} + V_{A\cap B\cap C\cap D} \right)}{V_D} \right] \quad \textbf{(38)}$$

In the analysis of complex mixtures there are others possible scenarios of overlapping peaks, and the general mathematical model for computing the volume of the overlapped region (V_{Oi}) of a peak *i* can be written as follows:

$$V_{O_i} = \sum_{i=1}^{n} V(p_i \cap p_{i+1}) + V(p_i \cap p_{i+2}) + \ldots + V(p_i \cap p_{i+n}) - V(p_i \cap p_{i+1} \cap p_{i+2}) - \ldots - V(p_i \cap p_{i+(n-1)} \cap p_{i+n}) + \ldots + (-1)^{n+1} V(p_i \cap p_{i+1} \cap \ldots \cap p_{i+n}) \quad \textbf{(39)}$$

where p_i is the peak under study, and p_{i+n} are the remaining overlapped peaks. The purity of the peak (P_i) is then given by Equation (40):

$$P_i = 1 - \left(\frac{V_{O_i}}{V_i} \right) \quad \textbf{(40)}$$

3.2.3. Global model of the two-dimensional chromatographic response function for application in complex mixtures

The definition of the new mathematical function for measuring and map the separation quality in LC × LC was performed on the basis of the objective function proposed by Duarte and Duarte (2010) for assessing the quality of LC of complex unknown samples (Equation (25)). The new CRF_{2D} is now expressed as:

$$CRF_{2D} = \sum_{i=1}^{N} P_i + N - f(t) \quad \textbf{(41)}$$

where $\sum P_i$ is the sum of all peak purity measures, N is the number of two-dimensional peaks, P_i is the purity of the peak i, and f (t) is a "time-economy" function. This economy function is related to the time needed to complete the two-dimensional separation process and it is given by the following equation:

$$f(t) = \left[\frac{(t_{R,L_{1D}} \times t_{R,L_{2D}}) - (t_{0_{1D}} \times t_{R,L_{2D}}) - (t_{0_{2D}} \times t_{R,L_{1D}}) + (t_{0_{1D}} \times t_{0_{2D}})}{(t_{R,L_{1D}} \times t_{R,L_{2D}})} \right] \quad \textbf{(42)}$$

where $t_{R,L_{1D}}$ and $t_{R,L_{2D}}$ are the retention times of the last eluted peaks in the first and second dimensions, respectively, and $t_{0_{1D}}$ and $t_{0_{2D}}$ are the elution times corresponding to the void volumes of the columns of the first and second dimensions, respectively. It should be further mentioned that when dealing with LC × LC analysis of complex organic mixtures,

the use of the number of two-dimensional peaks as a secondary criterion for mapping the separation quality is useful and highly desirable. In an ideal situation where all the two-dimensional peaks are well resolved, the maximum value of the sum of all peak purities equals the total number of peaks appearing in the chromatogram. However, in situations where the two-dimensional chromatogram exhibit overlapping peaks, the value of the sum of all peak purities is lower than the total number of peaks. This feature allows prospecting the global quality of the peaks (*Q*, in percentage), which informs about the degree of completion of the separation for a set of two-dimensional peaks:

$$Q(\%) = \left(\frac{\sum_{i=1}^{n} P_i}{N} \right) \times 100 \qquad \textbf{(43)}$$

However, it is important to highlight that *Q* is a relative measure of the quality of the peaks and not a good replacement for the quantification of the separation process, since a two-dimensional chromatogram can exhibit a *Q* close to 100% with a low number of well resolved peaks as compared to another LC × LC profile, which has a better overall separation quality.

3.3. Experimental

3.3.1. Reagents and solutions

All the chemicals used in this work were of analytical reagent grade and obtained from commercial suppliers without further purification. All the solutions were prepared with high purity water (18 M Ω cm).

Mobile phases for the LC × LC experiments were prepared with HPLC grade acetonitrile (ACN) and methanol (MeOH), ammonium acetate (CH_3COONH_4), acetic acid (CH_3COOH), di-sodium hydrogen phosphate dihydrate ($Na_2HPO_4.2H_2O$), and sodium di-hydrogen phosphate (NaH_2PO_4). The composition of the mobile phases for both

dimensions was adjusted according to the experimental conditions described in section 3.3.2. Prior to use, the mobile phases were filtered through membrane filters (PVDF, Gelman Sciences) of 0.22 μm pore size.

The experiments were conducted using a mixture of four aromatic compounds: syringic acid ($HOC_6H_2(OCH_3)_2CO_2H$, Mw = 198.17 $g.mol^{-1}$), ferulic acid ($HOC_6H_3(OCH_3)CH=CHCO_2H$, Mw = 194.18 $g.mol^{-1}$), caffeic acid ($(HO)_2C_6H_3CH=CHCO_2H$, Mw = 180.16 $g.mol^{-1}$), and 3-hydroxybenzoic acid ($HOC_6H_4CO_2H$, Mw = 138.12 $g.mol^{-1}$). The concentration of each solute in the samples was in the range of 0.235 – 0.300 $mg.mL^{-1}$, 0.215 – 0.369 $mg.mL^{-1}$, 0.482 – 0.669 $mg.mL^{-1}$ and 0.549 – 0.750 $mg.mL^{-1}$, respectively.

The samples were prepared by dissolving the appropriate amount of each compound in the mobile phase of the first dimension (section 3.3.2).

3.3.2. Instrumentation and chromatographic conditions

The first dimension consisted of a JASCO semi-micro HPLC pump (model PU-2085 Plus), a Rheodyne injection valve (model 7725i) equipped with a 20 μL loop, and an Acclaim Mixed-Mode HILIC-1 column (Dionex, Sunnyvale, CA, USA; diameter 4.6 mm; length 150 mm; comprised of 5 μm high-purity, porous, spherical silica particles with 120 Å diameter pores bonded with alkyl diol functional groups). The first dimension was operated in isocratic mode using different mobile phase compositions comprising of 100 mM CH_3COONH_4 / 0.3% (v/v) CH_3COOH (pH 5) and different amounts of ACN: 20, 40, 60, and 80% (v/v). The flow rate was 0.025 $mL.min^{-1}$ and the temperature of the analytical column was maintained at 30ºC in a JASCO column oven (model CO-2065 Plus). In the second dimension, a JASCO quaternary low pressure gradient pump (model PU-2089 Plus) and a reversed-phase Kromasil® 100-5-C18 column (Eka Chemicals AB - Separation Products, Bohus, Sweden; diameter 4.6 mm; length 150 mm; particle size 5 μm; pore diameter 100 Å) were applied. The second dimension was also operated in isocratic mode with a mobile phase composition consisting of 10mM phosphate buffer (pH 6) and 8%

(v/v) MeOH. The flow rate was 2.0 $mL.min^{-1}$ and the temperature of the analytical column was also maintained at 30ºC in a JASCO column oven. The outlet of the second dimension column was connected to a JASCO fluorescence detector (model FP-2020 Plus) operating at emission/excitation wavelengths of 240/410 nm.

The first and second dimensions were interfaced with an eight-port high pressure two-position interfacing valve (VICI® AG International, Schenkon, Switzerland) equipped with two identical 100 μL sampling loops. Modulation time was 240 s. The valve was controlled by the PSS WinGPC Unity software (Polymer Standards Service GmbH, Mainz, German) by receiving a start-up signal from a PSS Universal Data Center (model UDC 810).

The void volume of both first and second dimension columns was measured by injecting HPLC grade ACN under the same instrumental operation conditions and with a mobile phase composition of 100 mM CH_3COONH_4 / 0.3% (v/v) CH_3COOH (pH 5) and 60% (v/v) ACN. The retention time of the columns void volume (also termed void time) was determined by the first baseline disturbance caused by the elution of the organic solvent. The void times were set at 62.5 min and 0.6 min for the first and second dimension columns, respectively.

3.3.3. Software

The instrumentation was controlled and data set acquired in a PSS WinGPC Unity (Polymer Standards Service GmbH, Mainz, German) software. The algorithms developed in this study and data treatment were coded in MATLAB environment (The Mathworks Inc., Natick, MA, USA), in a computer with Intel® Core™2 Duo CPU and 4.00GB RAM memory.

3.4. Results and discussion

3.4.1. Application of the two-dimensional chromatographic response function to simulated chromatograms

The fitting and simulation of chromatograms is of great importance in the field of interpretative optimization procedures in LC. In the literature there are a variety of theoretical and empirical mathematical equations that describe the shape of one-dimensional chromatographic peaks (Di Marco and Bombi, 2001). The most simple mathematical peak function simulates the Gaussian elution profile which for two-dimensional applications can take the following form:

$$h_{(t_{1D},t_{2D})} = h_0 \times \exp\left[\left(\frac{t_{1D} - t_{R,1D}}{2\sigma_{1D}}\right)^2 + \left(\frac{t_{2D} - t_{R,2D}}{2\sigma_{2D}}\right)^2\right] \quad \textbf{(44)}$$

where the parameters appearing in this mathematical equation have the same meaning of those described in Equation (27). However, in both one-dimensional and two-dimensional chromatography practice, perfectly Gaussian peaks are not usually observed and most of them are characterized by a large asymmetry (either tailing or fronting). In this work, the rationale behind the simulation of two-dimensional chromatograms was not the prediction of the retention for a subsequent optimization study, but rather to assess the sensitivity of the developed CRF_{2D} (Equation (41)) towards different resolutions, number of two-dimensional peaks appearing in the chromatogram, and analysis time. For this purpose, a set of two-dimensional peaks were generated by the two-dimensional Gaussian function, expressed in Equation (44) and a two-dimensional Extreme Value function, expressed in Equation (28), and then summated to create simulated two-dimensional chromatograms with different peak arrangements. These functions were chosen taking into account the possible peak features in real experiments, ranging from the theoretically perfect Gaussian two-dimensional peaks to the most tailed peaks, generated by means of the Extreme Value function.

The two-dimensional chromatographic profiles depicted in Fig. III-10, aim at reproducing the possible situations obtained in real LC × LC experiments of complex samples. The separation quality of each simulated chromatogram was assessed by two different ways: a) using the peak purity concept developed in this study (Equation (40)); and b) the resolution measure (Equation (22)) proposed by Peters *et al.* (2007b). These two resolution criteria were used to compute the values of the CRF_{2D} (Equation. (41)) for each simulated chromatogram in order to evaluate and compare their performance as separation qualifiers of an entire two-dimensional chromatogram.

Table III-1 indicates, for each simulated chromatogram, the resolution values for each two-dimensional peak pair, the peak purity values for each two-dimensional peak, and the values of the CRF_{2D} estimated by means of the two resolution criteria. In both cases, the elution time corresponding to the column void volumes were set at 8 min for the first dimension and 0.5 min for the second dimension. As can be seen, the resolution metric of Peters *et al.* (2007b) has a larger number of failures than the new peak purity measure proposed in this study. As depicted in chromatograms **C** and **E** to **J** (Fig. III-10), there are peak pairs for which the detection of the saddle point fails, meaning that the resolution and, subsequently, the CRF_{2D} cannot be calculated for these two-dimensional chromatograms. Peak pair **1**-**3** in chromatogram **C** is an example: intuitively, one may consider that peaks **1** and **3** are neighbours and, therefore, it is expected to have a saddle point between these two peaks. However, according to the definition of Peters *et al.* (2007b) for peak-vicinity, peaks **1** and **3** are not considered to be "neighbours" since the trajectory profile connecting the peak maxima of these two peaks pass through the peak region of peak **2**. Therefore, computing the resolution between this two-dimensional peak pair is not considered to be meaningful because this peak pair fails the peak-vicinity test. Furthermore, even if the two-dimensional peaks **1** and **3** were recognized as "neighbours", the interference of peak **2** completely distort the surface region between those two peaks, which hinders the detection of a saddle point. On the other hand, for chromatograms **A**, **B** and **D**, all peaks pass the peak-vicinity test and, therefore, the resolution (Equation (22)) can be calculated between them, allowing to compute and to assess the overall quality of the two-dimensional chromatographic separation. On this regard, the CRF_{2D} considers chromatograms **A** and **B** of less quality than chromatogram **D**, being this a direct consequence of their lower number of peaks.

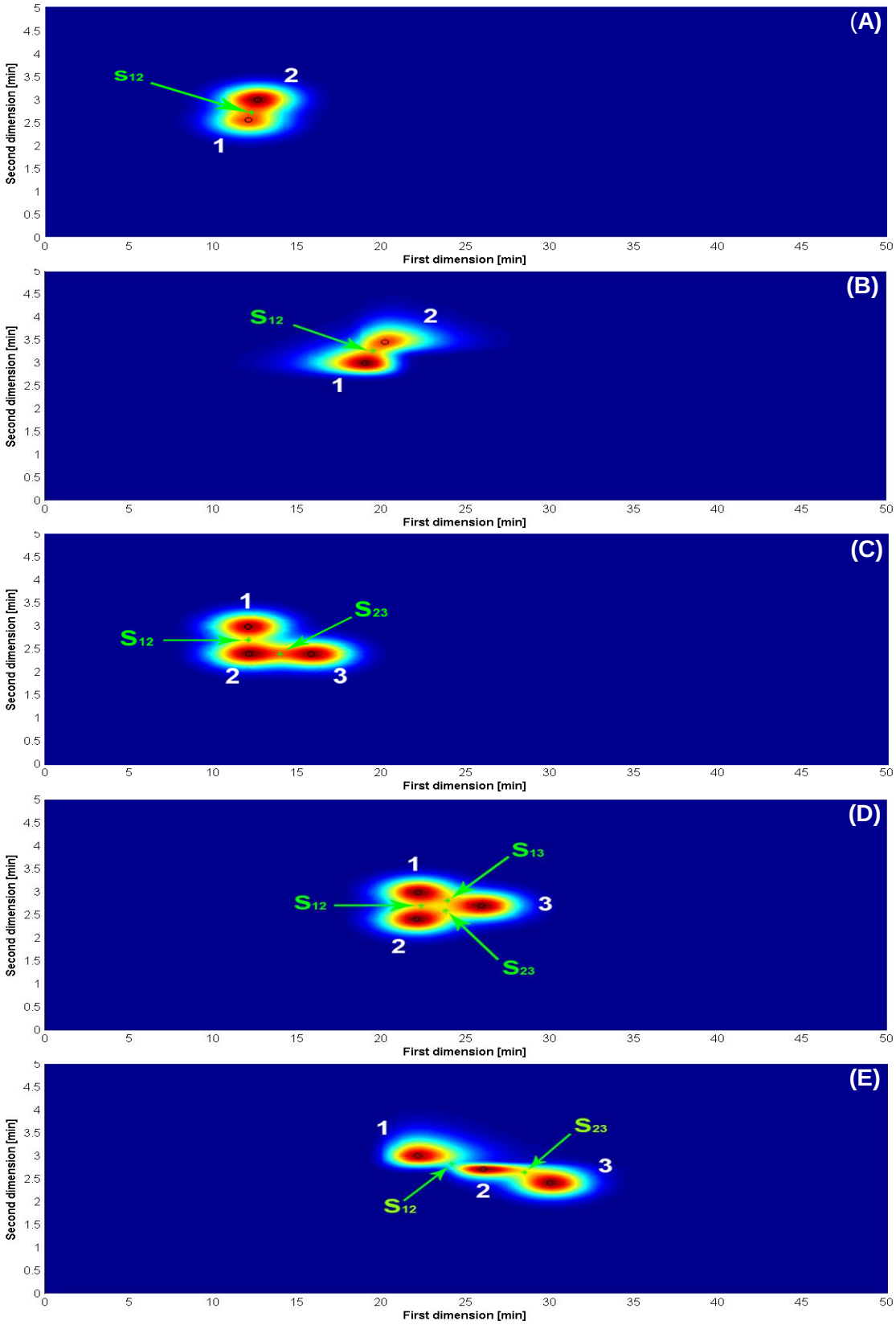

Fig. III-10. Simulated chromatograms used to demonstrate the performance of the developed CRF_{2D}. $S_{i,j}$ refers to the saddle point between peaks *i* and *j*. Refer to Table III-1 for further information on these simulations.

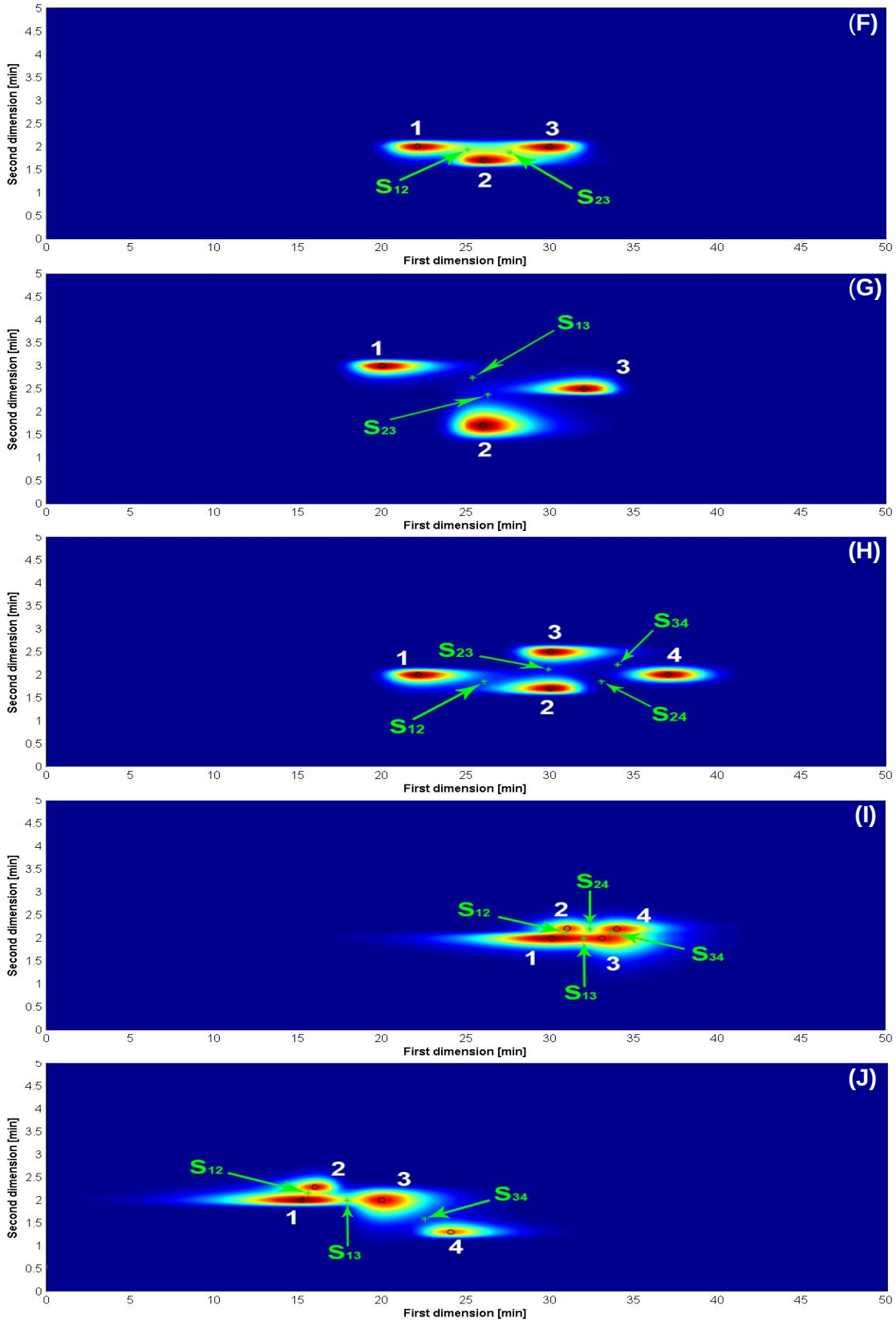

Fig.III-10 Simulated chromatograms used to demonstrate the performance of the developed CRF_{2D}. $S_{i,j}$ refers to the saddle point between peaks *i* and *j*. Refer to Table III-1 for further information on these simulations. (cont.)

Table III-1. Resolutions, purity and response function values calculated for the ten simulated chromatograms shown in Fig. III-10.

Simulation	Resolution (Rs)[(a)]						Peak purity (P_i) [(b)]				CRF_{2D} (Rs) [(c)]	CRF_{2D} (P_i) [(d)]
	1-2	1-3	1-4	2-3	2-4	3-4	1	2	3	4		
A	0.65	-	-	-	-	-	0.78	0.82	-	-	2.3	3.3
B	0.70	-	-	-	-	-	0.88	0.83	-	-	2.2	3.2
C	0.75	*FAIL*	-	0.68	-	-	0.86	0.73	0.81	-	*FAIL*	5.0
D	0.74	0.74	-	0.75	-	-	0.79	0.80	0.80	-	4.7	4.8
E	0.90	*FAIL*	-	0.75	-	-	0.96	0.67	0.84	-	*FAIL*	4.9
F	0.82	*FAIL*	-	0.81	-	-	0.82	0.66	0.70	-	*FAIL*	4.6
G	*FAIL*	1.4	-	1.2	-	-	1.0	0.98	0.96	-	*FAIL*	5.3
H	1.1	FAIL	*FAIL*	1.2	1.54	1.4	0.92	0.90	0.98	1.0	*FAIL*	7.2
I	0.66	0.63	*FAIL*	*FAIL*	0.76	0.65	0.85	0.81	0.73	0.63	*FAIL*	6.4
J	0.76	0.77	*FAIL*	*FAIL*	*FAIL*	1.1	0.85	0.68	0.91	0.91	*FAIL*	6.8

FAIL Saddle point not detected.

(a) Resolution measure (R_s, Equation (22)) of Peters *et al.* (2007b) using the valley-to-peak ratio and the saddle point concept.

(b) Peak purity (P_i) calculated through Equation (40)).

(c) CRF_{2D} calculated through Equation (41) using the resolution measure (R_s, Equation (22)) of Peters *et al.* (2007b) ($t_{0\,1D}$ = 8min; $t_{0\,2D}$ = 0.5min).

(d) CRF_{2D} calculated through Equation (41) using the peak purity (P_i) concept of Equation (40) ($t_{0\,1D}$ = 8min; $t_{0\,2D}$ = 0.5min).

When using the two-dimensional peak purity concept (Equation (40)) as the quality separation criterion, the CRF_{2D} function also classifies **A** and **B** as the less desirable chromatograms, which again reflects the importance of the number of distinguishable two-dimensional peaks as a secondary requirement for ranking the quality of the chromatographic separations. Chromatograms **C** to **G** all exhibit the same number of two-dimensional peaks. However, and due to the emphasis that it places on the purity of each peak, chromatogram **F** with a high degree of peak overlapping is considered of less quality than chromatogram **G**, which exhibits peak purities close to 1.0. The order by which the CRF_{2D} rank the quality of these chromatograms depends not only on the purity of each peak, but also on the total time of analysis. For example, chromatograms **C** and **D** exhibit similar overall resolution (ΣP_i is around 2.4). However, chromatogram **C** elute in a smaller time window, which contribute to rank this chromatogram as a more desirable chromatogram. The importance of the peak purity term is also reflected in the values of the CRF_{2D} of chromatograms **H**, **I** and **J**; although exhibiting the same number of peaks, values of peak purity close to 1.0 in chromatogram **H** contributes to rank this chromatogram as the most desirable one.

3.4.2. Application of the two-dimensional chromatographic response function for assessing the separation quality of complex mixtures

The validity of the new CRF_{2D} for qualifying the separation degree attained under different LC × LC conditions was further assessed through a set of experiments with a mixture of four aromatic compounds, and using the amount of ACN in the mobile phase of the first dimension as the experimental variable (see section 3.3.2 for additional details on the chromatographic conditions). It must be emphasized, however, that the objective of this study was not to separate these organic compounds by polarity and/or hydrophobicity, but to demonstrate the efficiency of the new CRF_{2D} for measuring the separation quality of real LC × LC. These experiments will also provide good indications on the feasibility of the CRF_{2D} as a valuable tool for the design of suitable LC × LC procedures.

Before assessing the overall separation quality of each two-dimensional chromatogram, it is necessary to perform some data pre-treatment in order to reduce the effects of the experimental variations. Firstly, the experimental data points in the first dimension of the two-dimensional matrix were interpolated using a spline algorithm. Secondly, this same algorithm was applied to the experimental data points of the second dimension, but in this case to remove data points in order to smooth the experimental data set and to generate a square grid with spacing as close as possible of that used in the first dimension. Thirdly, spurious data points caused by low levels of random background noise were removed by considering a threshold for the minimal peak height acceptability. In this study, a peak was accepted as such when its height was higher than 10% of the height of the most intense peak in the chromatogram. After these data pre-treatment procedures, the methodology for two-dimensional peak detection and CRF_{2D} calculation followed closely the steps described in section 3.2.

Fig. III-11 depicts the regions of interest of each two-dimensional chromatogram of the mixture of four aromatic compounds obtained for the different amounts of ACN in the mobile phase of the first dimension. Before going any deeper into the application of the algorithm for peak detection suggested in section 3.2.1, a visual inspection of the two-dimensional chromatograms in Fig. III-11A, Fig. III-11C, and Fig. III-11D indicates a

number of peaks higher than would be expected from the analysis of the mixture of four organic compounds. These results are likely to be a consequence of the occurrence of unpredictable retention mechanisms (e.g. hydrophobic interactions) of the solutes into the column packing material. The apparent high number of peaks as a result of peak broadening in the time window between 2.5 and 5 minutes in the second dimension is a clear example of such phenomena, i.e. strong adsorption between the solute (in this case, ferrulic acid) and the column stationary phase (reversed-phase C18). In light of these results, it is recommended to devote a special attention to the optimization of the chromatographic conditions in order to minimize the occurrence of unwanted interactions between the stationary phase responsible for poor peak shape and unpredictable retention, and the analytes.

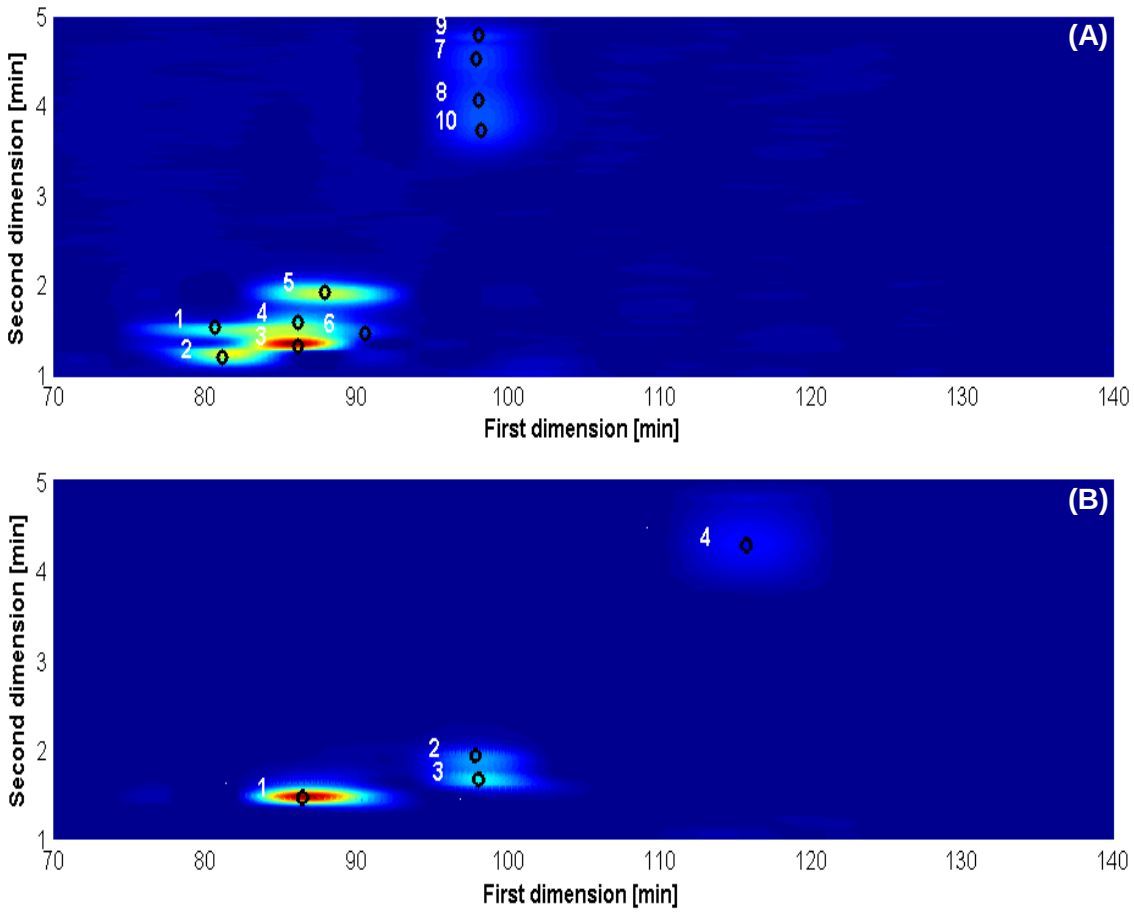

Fig. III-11. Two-dimensional chromatograms of mixtures of four aromatic compounds obtained with different amounts of ACN in the mobile phase of the first dimension: 20% (**A**), 40% (**B**), 60% (**C**) and 80% (**D**). The retention time axes were expanded for a better visualization of each 2D chromatogram, and the detected peaks are identified by a number.

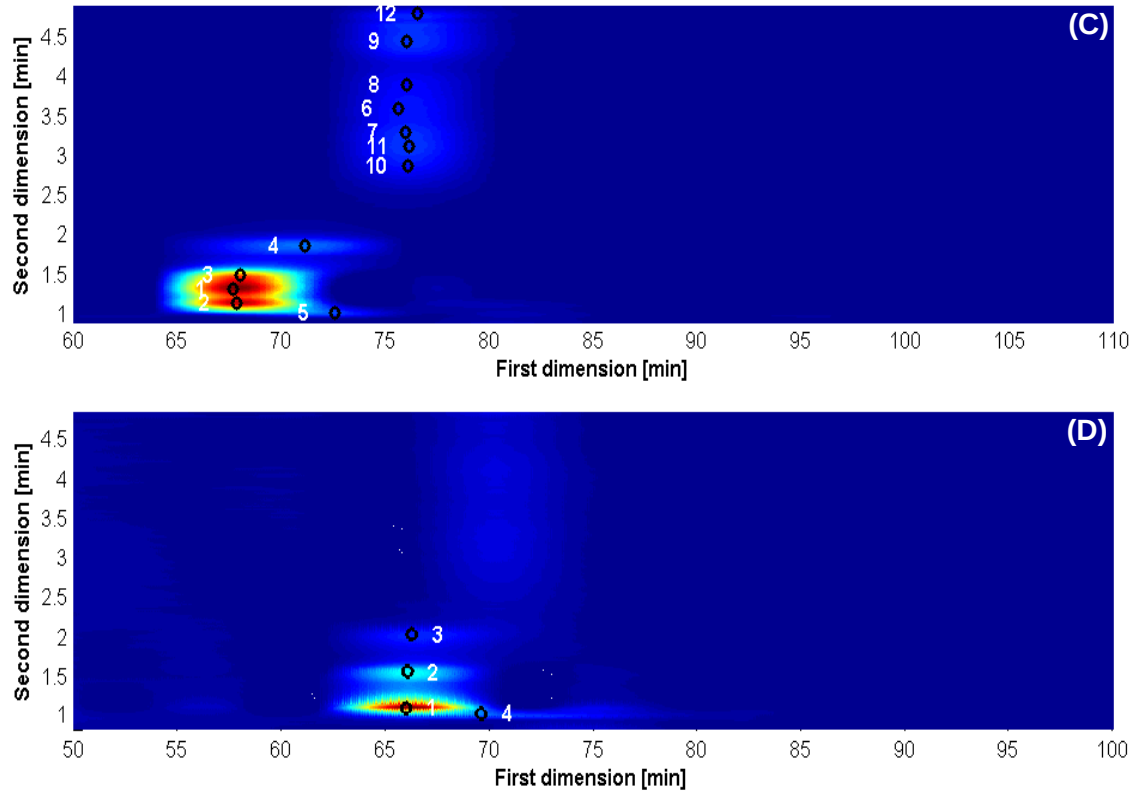

Fig.III-11. Two-dimensional chromatograms of mixtures of four aromatic compounds obtained with different amounts of ACN in the mobile phase of the first dimension: 20% (**A**), 40% (**B**), 60% (**C**) and 80% (**D**). The retention time axes were expanded for a better visualization of each 2D chromatogram, and the detected peaks are identified by a number. (cont.)

At the completion of the peak detection algorithm, 10 peaks have been detected in the two-dimensional chromatogram of Fig. III-11A. These peaks were divided into two clusters of overlapping peaks: the cluster containing the peaks **1** to **6**, and the cluster with peaks **7** to **10**. The time windows of the chromatogram containing these two clusters were fitted using a mathematical model of two-dimensional Log-normal functions (Equation (27)). The quality of the fit procedure is supported by the values of the R^2_{adj}: 0.9186 and 0.9696, respectively.

In the chromatogram of Fig. III-11B, 4 peaks were recognized using the proposed algorithm: peaks **1** to **3** are overlapped, while peak 4 is isolated with no surrounding peaks. The time window containing peaks **1** to **3** was fitted with two-dimensional Log-normal functions, resulting in a R^2_{adj} of 0.9383. In the chromatogram Fig. III-11C, the peak detection algorithm was able to detect 12 peaks, which were divided into two clusters: one cluster containing peaks **1** to **5**, and a second cluster containing peaks **6** to **12**.

The R^2_{adj} for the fit of the time windows of these two clusters were 0.9805 and 0.9699, respectively. Finally, the peak detection algorithm has identified only 4 peaks, all overlapped in one cluster, in the chromatogram depicted in Fig. III-11D. The fitting procedure of the time window of this cluster exhibited an R^2_{adj} of 0.9060. The adequacy of the fitted mathematical model in chromatogram **D** seems to be relatively low. This result is likely to be a consequence of the low level of the detector response (below the stipulated 10% threshold), which is confounded with the background noise of the two-dimensional chromatogram.

Nevertheless, these results highlight the success of both peak detection algorithm and fitting procedure in determining the coordinates of the two-dimensional peaks automatically from data obtained from real experimental two-dimensional chromatograms. These results also illustrate the importance of selecting appropriate peak detection thresholds and also the difficulties in identifying the two-dimensional peaks just by visual inspection of the chromatograms.

Once the peaks have been detected and fitted in each two-dimensional chromatogram, the purity of each peak can be computed, and the assessment of the overall separation quality using the new CRF_{2D} can be completed. The values of the peak purities and of the CRF_{2D} obtained for each experiment are shown in Table III-2. It can be observed that the highest values of the CRF_{2D} were obtained for experiments **A** and **C**, whereas the lowest CRF_{2D} value was obtained for experiment **D**. The reason why the new CRF_{2D} ranks chromatogram **C** as the most desirable one in comparison with chromatogram **A**, relies on the number of detected peaks, which are higher for chromatogram **C**. In fact, the time needed to complete the analysis is similar in both chromatograms **A** and **C**, as well as the global quality of the peaks (*Q* is equal to 52 and 53% for chromatograms **A** and **C**, respectively).

On the other hand, chromatograms **B** and **D** exhibit the same the number of detected peaks, although the CRF_{2D} classifies chromatogram **D** as being of less quality. This result is likely to be explained by the lower chromatographic separation degree exhibited by chromatogram **D**, being this reflected in the values of the global quality of the peaks (*Q* is equal to 89% and 67% for chromatograms **B** and **D**, respectively).

These results demonstrate that the new CRF_{2D} efficiently discriminate between chromatograms with different separation quality, in terms of the number of two-dimensional peaks and degree of separation reached for each two-dimensional peak. Nevertheless, and disregarding the fact that one is dealing with a mixture of well-known compounds, the qualitative grading of the two-dimensional chromatograms using the CRF_{2D} (Table III-2) parallels the chromatographic data, i.e., the CRF_{2D} values are in line with the intuitive choice of an analyst. It should be mentioned, however, that when dealing with complex mixtures of organic compounds, the advantage of using a CRF_{2D} for measuring and map the quality of the two-dimensional chromatographic separation only holds if a careful attention is devoted to development of reliable chromatographic conditions, including the selection of appropriate columns in both dimensions (i.e. type of stationary phases and column sizes), mobile phases (i.e. organic modifier and buffer), elution modes (i.e. isocratic or gradient elution), flow rates, and volume of the transferred fractions and frequency of the sample transfer between the first and second dimension.

It should also be mentioned that the time-economy function (Equation (42)) does not exert a major influence on the values of the CRF_{2D}, since it can only discriminate between chromatograms bearing the same number of peaks and with similar peak purity values. In the case of the two-dimensional chromatograms depicted in Fig. III-11, the only situation where the time-economy function had some effect on the values of the CRF_{2D} was when comparing experiments **B** and **D**, but though not enough to change the qualitative grading of these two-dimensional chromatograms on the basis of the CRF_{2D} values.

Table III-2. Individual peak purity and CRF_{2D} values for the four two-dimensional chromatograms presented in Fig. III-11.

Experiment	Peak purity (P_i) [(a)]												CRF_{2D} [(b)]
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0.620	0.816	0.647	0.391	0.760	0.433	0.328	0.293	0.479	0.421	-	-	14.9
B	0.995	0.728	0.855	1.00	-	-	-	-	-	-	-	-	7.18
C	0.538	0.728	0.693	0.941	0.665	0.606	0.315	0.430	0.491	0.542	0.033	0.345	18.2
D	0.922	0.878	0.785	0.106	-	-	-	-	-	-	-	-	6.62

(a) Peak purity (P_i) calculated through Equation (40)).

(b) CRF_{2D} calculated through Equation (41) using the peak purity (P_i) concept of Equation (40).

3.5. Conclusions

A new CRF_{2D} has been proposed for the estimation of the quality index of separation in LC × LC of complex organic mixtures. This function has evolved from the objective function proposed by Duarte and Duarte (2010) for assessing the quality of LC of complex unknown samples. The new CRF_{2D} also takes into account the most important criteria for measuring and rank the separation quality of any chromatographic procedure: overall resolution of the two-dimensional chromatogram, number of distinguishable two-dimensional peaks in the chromatogram, and time needed to complete the analysis. One important outcome of this study was the development of an algorithm for detecting two-dimensional peaks automatically from real experimental LC × LC data, followed by the application of a mathematical model to fit the obtained chromatographic data. The other important problem tackled in this study was the estimation of the quality of the two-dimensional chromatographic separation for complex organic mixtures, which was accomplished by adapting the peak purity concept to two-dimensional separations. The use of this concept changed the paradigm of the classical resolution metric for LC × LC, which is now oriented to each individual two-dimensional peak instead of each two-dimensional peak pair. Furthermore, this two-dimensional peak purity concept is a normalized measure that does not require the prior definition of a minimum acceptable threshold, thus facilitating the combination of each elementary two-dimensional peak purity into a single global value. The theoretical study with simulated two-dimensional chromatograms indicated that the CRF_{2D} places more emphasis on the purity of each two-dimensional peak over total time of analysis. The number of distinguishable two-dimensional peaks was also found to be an important secondary requirement for ranking the quality of the chromatographic separations.

The validity of the new function for qualifying the two-dimensional separation degree attained under different chromatographic conditions was further assessed using a mixture of four organic compounds. It was demonstrated that the new CRF_{2D} gives indicative values for discriminating between two-dimensional chromatograms with different separation quality. Nevertheless, the use of an algorithm that detects all two-dimensional peaks may provide false information about the sample, which means that an

adjustable threshold must be defined for the peak detection. The results from this study also provide information on the importance of developing optimum chromatographic conditions in order to minimize the occurrence of unpredictable retention phenomena between the analytes and the column packing material.

IV

Application of a two-dimensional chromatographic response function for red wine fingerprinting

4.1. Introduction

In the past decades, many chemometric methods have been developed to perform a chemical characterization of food products. This characterization is intended to classify and authenticate these products according to their sources, quality, variety, and type of products (Moret *et al.*, 1994). Within the wide range of products amenable to authentication, wine has been extensively studied because it is a product manufactured and consumed world-wide that can be easily adulterated (Arvanitoyannis *et al.*, 1999). From a chemical point of view, wine is viewed as a complex matrix composed by different constituents such as water, sugars, ethanol, glycerol, organic acids and salts in higher concentrations, as well as aliphatic and aromatic alcohols, amino acids and phenolic compounds, in lower concentrations (Tarantilis *et al.*, 2008). Therefore, the development of techniques for the characterization of wine samples according to several features that conditioning the specific and individual wine composition, such as the region of production, grape variety, climate, yeasts, viticultural practices, processing methods, and vintage, is extremely important to assure the quality control, the authenticity, and geographical certification both for winemaking industries and the final customers (Rodrigues *et al.*, 2009, Tarantilis *et al.*, 2008).

In the past decades, especially after joining the European Union, Portugal has one of the largest wine markets and is one of the most important wine-exporting countries in the world. The most important wine produced is the worldwide known "Vinho do Porto" or "Port wine". Since the early times, Portugal gave importance to the wine forgery problems. To account for these problems, the world's first demarcated wine region, the Douro, was created in 1756, alongside an agency to control each step of the production in order to protect the Port wine from producers and shippers that began enlarging the supply to England by adulterating the wine with inferior quality grapes. Since that time, other demarcated regions were created, and nowadays nineteen Controlled Denomination of Origin (DOC) regions exist, where wines with high quality standards and with some very specific and controlled features are produced (Henderson and Rex, 2011).

During this last decade, several analytical methods have been developed with the purpose of performing a chemical and biochemical characterization of different wine samples (Airado-Rodríguez *et al.*, 2009, Chira *et al.*, 2011, Falcão *et al.*, 2008, Gómez-Alonso *et al.*, 2007, Guadalupe *et al.*, 2012, Marcobal *et al.*, 2005, mo Dugo *et al.*, 2006, Santos *et al.*, 2004, Šeruga *et al.*, 2011, Weldegergis *et al.*, 2011). However, due to the complexity of wine, a fully characterization of all chemical components is very difficult to achieve and it requires an extremely time-consuming analysis procedure. To address this problem, the present-day trends in wine analysis focus on the development of techniques capable of characterizing the sample according to their chemical pattern with the purpose of identifying a fingerprint of the sample. This new approach is feasible since for each plant variety, the wide range of chemical components such as anthocyanins, flavonoids, procyanidins, and hydroxycinnamic acids, generates a characteristic fingerprint (Tarantilis *et al.*, 2008). However, it is important to acknowledge that some external variables, such as the weather pattern or the different processes of vinification of each winemaking industry, can hinder a chemical pattern recognition (Rodrigues *et al.*, 2009). Nowadays, the association of multichannel detectors to multi-dimensional separation techniques, especially LC × LC, constitutes a huge potential for accomplishing the chemical pattern recognition of a wine sample. Until recently, one of the major drawbacks that hindered the application of these analytical techniques was the mathematical procedures for data treatment. Nevertheless, the recent developments within this field, specially made by the Rutan's research group (Allen and Rutan, 2011, Bailey and Rutan, 2011, Bailey *et al.*, 2011), allows for several of these problems being already cracked. Another drawback is the massive time spent in just one analysis: typical LC × LC run times could be of hours or even days (Porter *et al.*, 2006). However, Stoll *et al.* (2006) have designed a LC × LC system with a reduction of analysis time of 15 to 30 min. Although not widespread, is expected that this new design will increase the use of LC × LC technique in the near future. It should also be mentioned, that the LC × LC technique has already been used for the characterization of components of red wine samples (Cacciola *et al.*, 2007, Dugo *et al.*, 2009), although no publication has been found in the literature using this technique to originate a wine fingerprint.

Therefore, in this work, three commercial Portuguese red wine samples were analysed using a comprehensive two-dimensional liquid chromatographic system coupled

to a diode array and fluorescence detectors (LC × LC–DAD–FLD) with the aim of creating a two-dimensional chromatographic wine fingerprinting of the samples. For this purpose, the CRF_{2D} developed in Chapter III is applied for quantifying a so-called "chromatographic value" of the wine fingerprint. Each wine sample was also characterized by means of excitation-emission matrix fluorescence spectroscopy in order to identify the specific excitation/emission wavelengths to be employed in the subsequent LC × LC–DAD–FLD analyses.

4.2. Experimental

4.2.1. Wine samples

Three commercial red wine samples were chosen for this study. These samples are originated from three denominated Portuguese wine regions: Douro, Alentejo, and Dão. The geographic locations of these regions are represented in Fig. IV-1. Table IV-1 describes the most relevant information of each sample: the origin region, the type of grapes, and the year of production. Additional information regarding each sample can be found in Annex ii. Bottles containing each wine sample were opened, aliquots of ca. 15 mL were sampled from each bottle and protected against sunlight and stored at 4°C until further analysis.

Table IV-1. Information regarding each studied wine sample.

Sample	Region	Grapes	Vintage
A	Douro	• Aragonês • Touriga Nacional • Touriga Franca • Tinta Barroca	2008
B	Alentejo	• Aragonês • Alfrocheiro • Castelões	2008
C	Dão	• Aragonês • Touriga Nacional • Alfrocheiro • Jaen	2008

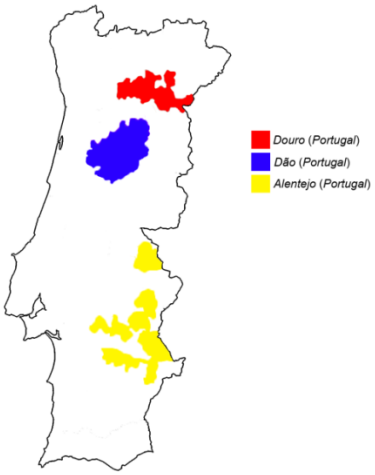

Fig. IV-1. Schematic representation of mainland Portugal with the Douro (red), Dão (blue), and Alentejo (yellow) wine regions.

4.2.2. Reagents and solutions

All the chemicals used in this study are already described in section 3.3.1. Please, see this section for additional details.

4.2.3. Excitation-emission matrix fluorescence spectroscopy of wine samples

The fluorescence spectra were recorded on a spectrophotometer JASCO, model FP-6500. The EEM fluorescence spectroscopy involved scanning and recording of 18 individual emission spectra (230-500 nm) at sequential increments of 10 nm of excitation wavelength between 220 and 400 nm. The spectra were recorded at a scan speed of 100 nm min^{-1} using excitation and emission slit bandwidths of 10 nm. Each wine sample was diluted in a concentration of 0,08% of sample to 99,93% of mobile phase (v/v) applied in the LC × LC–DAD–FLD analyses (section 4.2.4) to avoid fluorescence signal saturation. The peaks due to water Raman scatter were eliminated from all spectra by subtracting the mobile phase blank spectra.

4.2.4. Instrumentation and chromatographic conditions

The two-dimensional chromatographic system is the same used in section 3.3.2 (Chapter III), with some modifications. In the first dimension, the chromatographic column was replaced for an Acclaim Mixed-Mode WAX-1 column (Dionex, Sunnyvale, CA, USA; diameter 4.6 mm; length 150 mm; comprised of 5 μm high-purity, porous, spherical silica particles with 120 Å diameter pores bonded with alkyl amine functional groups). The first dimension was operated in isocratic mode using a mobile phase composition containing 10mM phosphate buffer (pH 3) and 1% (v/v) ACN. The flow rate was 0.080 $mL.min^{-1}$ and the temperature of the analytical column was maintained at 50ºC. The second dimension was also operated in isocratic mode with the same mobile phase composition of the first dimension. The flow rate was 3.0 $mL.min^{-1}$ and the temperature of the analytical column was also maintained at 50ºC. Modulation time was 85s. The outlet of the second dimension column was connected to two detectors, in series: a DAD (JASCO, model MD-2010) operating between 201 and 601 nm, and a FLD (JASCO, model FP-2020 Plus) operating at emission/excitation wavelengths of 230/320 nm.

Prior to each chromatographic analysis, the wine samples were filtered through HPLC Certified Syringe Filters (SPARTAN, Whatman GmbH, Germany) of 0.20 μm pore size. Afterwards, the sample was diluted in 20% of the mobile phase (v/v) to prevent changes in pH along the analysis.

4.2.5. Software

The software used in this chapter is the same as mentioned in section 3.3.3, except for DAD data acquisition, where the *ChromNav* Chromatography Data System (JASCO corporation, Tokyo, Japan) was used instead.

4.3. Results and discussion

4.3.1. Excitation-emission matrix fluorescence spectroscopy

The EEM fluorescence spectroscopy has already been used for wine fingerprinting (Airado-Rodríguez *et al.*, 2009, Yin *et al.*, 2009). In fact, wine samples contain several molecules with intense fluorescence properties (e.g. phenolic acids, anthocyanins, flavanols, and tannins), and the type and the amounts of these fluorescents compounds can change according to the variety of grapes, wine processing or ageing.

Table IV-2 describes, for each wine sample, the peak location in terms of emission and excitation wavelengths and their respective fluorescence intensity, whereas Fig. IV-2, Fig. IV-3 and Fig. IV-4 shows the EEM spectrum of each sample.

Table IV-2. Emission and excitation wavelengths of each identified peak in the wine samples and their respective fluorescence intensity.

Sample	Peak	Excitation Wavelength (nm)	Emission Wavelength (nm)	Fluorescence Intensity (AU)
A	1	230	316	545.60
	2	270	364	435.60
	3	280	318	470.44
B	1	230	322	611.88
	2	270	364	529.15
	3	280	334	490.55
C	1	230	316	600.98
	2	230	362	498.68
	3	260	380	651.06
	4	280	316	516.93

In sample *A*, three fluorescent peaks corresponding to three different excitation/emission wavelength pairs are discernible: 230/316 nm, 270/264 nm, and 280/318 nm. On the other hand, sample *B* has a very similar spectral pattern, exhibiting the same 270/364 nm pair, whereas in the other two pairs only the emission wavelength suffers a shift towards higher values: 230/322 nm and 280/334 nm, respectively. Sample *B* also exhibits higher fluorescence intensity than sample *A*, which suggests the presence of more

fluorescent compounds in the former sample. In the wine samples, the peaks with higher fluorescence intensity are located at 230/316 nm and 230/322 nm for sample *A* and *B*, respectively.

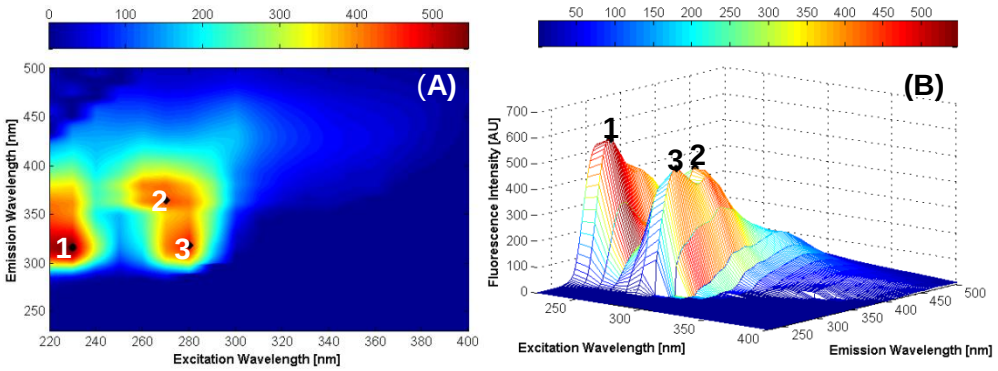

Fig. IV-2. Contour map (**A**) and three-dimensional graph (**B**) representation of the EEM fluorescence spectrum of sample *A*.

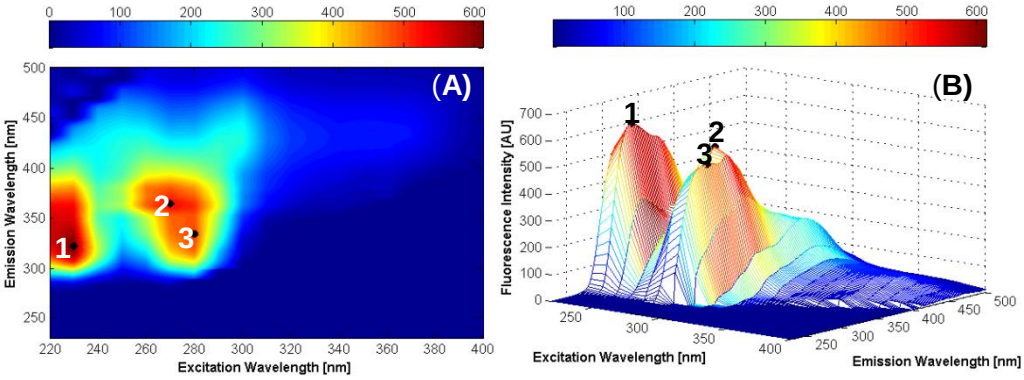

Fig. IV-3. Contour map (**A**) and three-dimensional graph (**B**) representation of the EEM fluorescence spectrum of sample *B*.

On the other hand, sample *C* shows four fluorescent peaks at the following excitation/emission wavelength pairs: 230/362 nm, 260/380 nm, 280/316 nm, and 230/316 nm pair (at the same wavelengths of peak **1** in sample *A*). The most intense peaks are peak **3**, at excitation/emission wavelength of 260/380 nm, immediately followed by peak **1**, at excitation/emission wavelength of 230/316 nm.

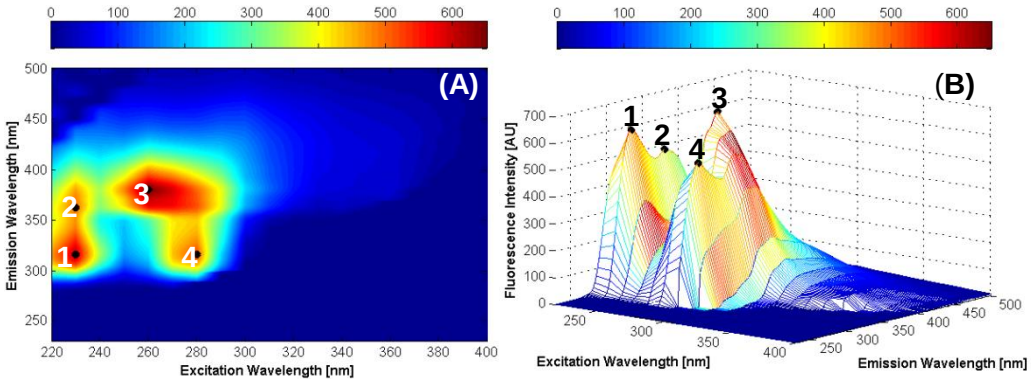

Fig. IV-4. Contour map (**A**) and three-dimensional graph (**B**) representation of the EEM fluorescence spectrum of sample *C*.

Overall, since for all spectra the higher fluorescence values were found in the region of peak **1**, particularly in the spectra of samples *A* and *B*, the excitation/emission wavelength pair of 230/320 nm was chosen for performing the subsequent LC × LC–DAD–FLD analyses.

4.3.2. Two-dimensional liquid chromatography analysis

4.3.2.1. Data pre-treatment

After obtaining the LC × LC experimental data, and before data analyses, it is necessary to modulate the data and ensure a reduction of some undesired variations. To resolve some of these issues a data pre-treatment algorithm has been designed. This algorithm starts by generating the two-dimensional chromatographic matrix in a previously chosen wavelength, for the case of DAD. To accomplish this, the algorithm starts by dividing the vector data for the chosen wavelength in the corresponding second dimension slices according to the modulation time of the valve. Afterwards, from a visual inspection of the data, three major problems have been found: a) the existence of chromatographic information in the void time in both dimensions, locations where the baseline instability inherent to the chromatographic process exists, b) non-stable background, and c) some experimental noise. To solve these problems, the algorithm starts by correcting (if

necessary) the slices correspondent to the void time in the first dimension and the existence of values below the baseline in each slice (negative peaks). Subsequently, the minimum value of each slice is subtracted to each slice value for correcting the two-dimensional background. The third step of the algorithm entails the correction of the instability present in each slice, which refers to the void time in the second dimension. To accomplish this, a slice with this variation, created using the minimal chromatographic point from all slices, is subtracted to each slice. Then, to remove background noise, a smoothing algorithm with a moving average method is applied to all slices. However, since only peaks above a previously selected threshold are identified, an algorithm replaces all slices in which its maximum value does not reach 5% of the most intense peak in the two-dimensional chromatogram, by a slice of zeros. In the remaining slices, all values that do not reach 3% of the most intense peak in the two-dimensional chromatogram are replaced by zero. All points appearing below the void time of the second dimension in each slice are replaced by zeros too. Finally, the two-dimensional matrix is created and the data set are normalized.

4.3.2.2. Two-dimensional chromatographic profiles based on the diode array detector

The DAD gives the opportunity to study the chromatographic profile of the sample at various wavelengths between 201 and 601 nm. As already explained in section 2.2.1.2 (Chapter II), it is impossible to plot a three-dimensional data matrix and, in this case, is necessary to build an individual two-dimensional chromatogram for each DAD wavelength. Consequently, it becomes necessary to choose an appropriate wavelength for wine fingerprinting. In this study, the selection of the wavelength was made based on the information present in each two-dimensional chromatogram, plus the background noise. Fig. IV-5 illustrates the two-dimensional chromatogram of the three wine samples at 279 nm, which was found to provide the best signal-to-noise response (in Annexes iii, iv and v five additional chromatograms of the same samples at different wavelengths can be found). It should be mentioned, however, that the remaining spectral information can be used to confirm, for instance, the existence of overlapping peaks. To demonstrate this one can perform a simple exercise with the LC × LC-DAD data of sample *B*. Fig. IV-6 shows eight

two-dimensional chromatographic sections, between 21.25 and 48.75 min in the first dimension and 0.25 and 0.65 min in the second dimension, at different wavelengths. For each case, the two-dimensional peaks are detected using the sum of the second-order derivatives and constrained intensities above the threshold of 7.5% of the peak with highest intensity in the two-dimensional chromatogram. It is possible to verify that additional information on the chromatographic profile of the sample and, therefore, on its chemical pattern, can be derived using different wavelengths.

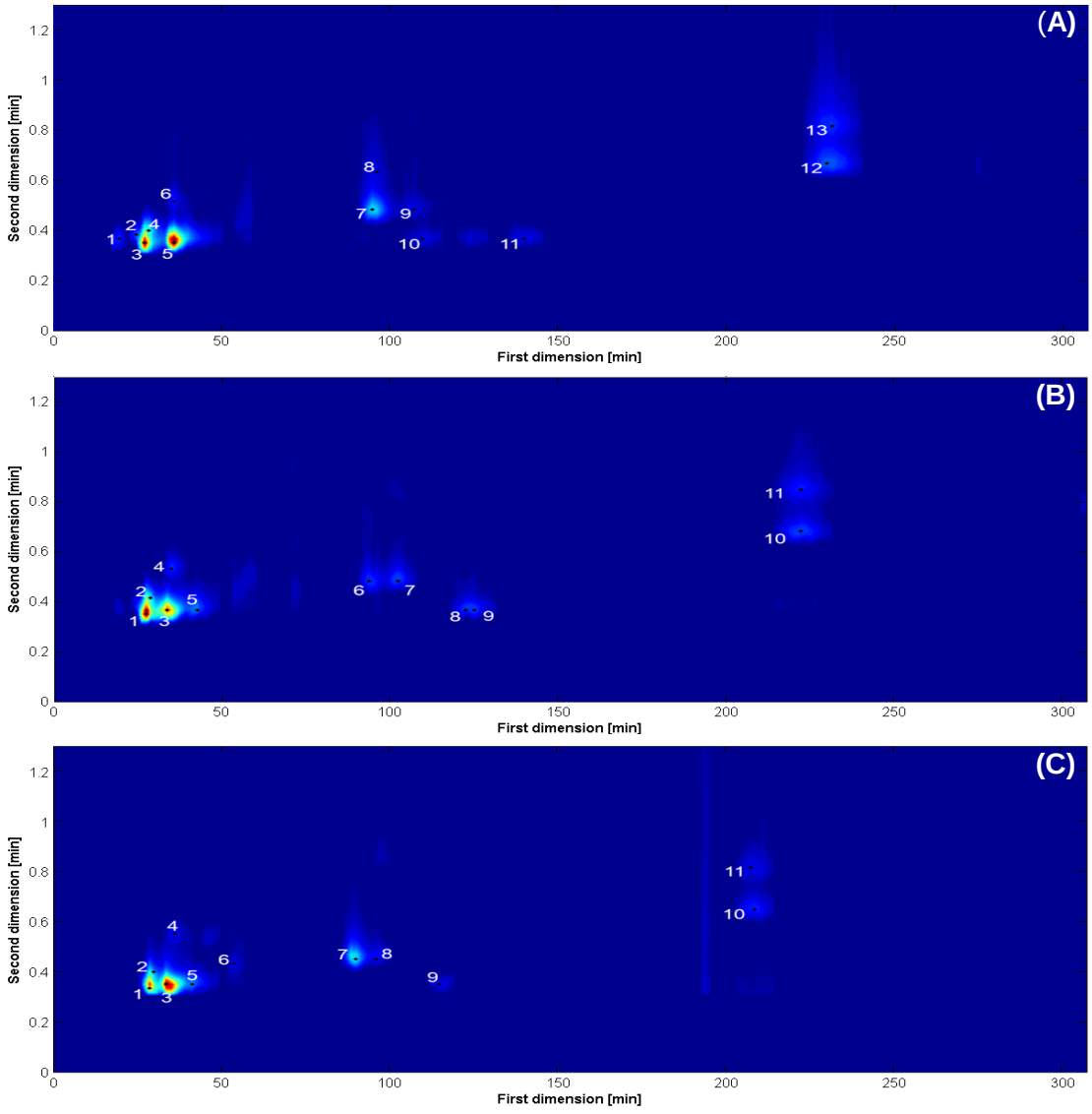

Fig. IV-5. Representation of the LC × LC-DAD at 279 nm for the wine sample *A* (**A**), wine sample *B* (**B**), and wine sample *C* (**C**).

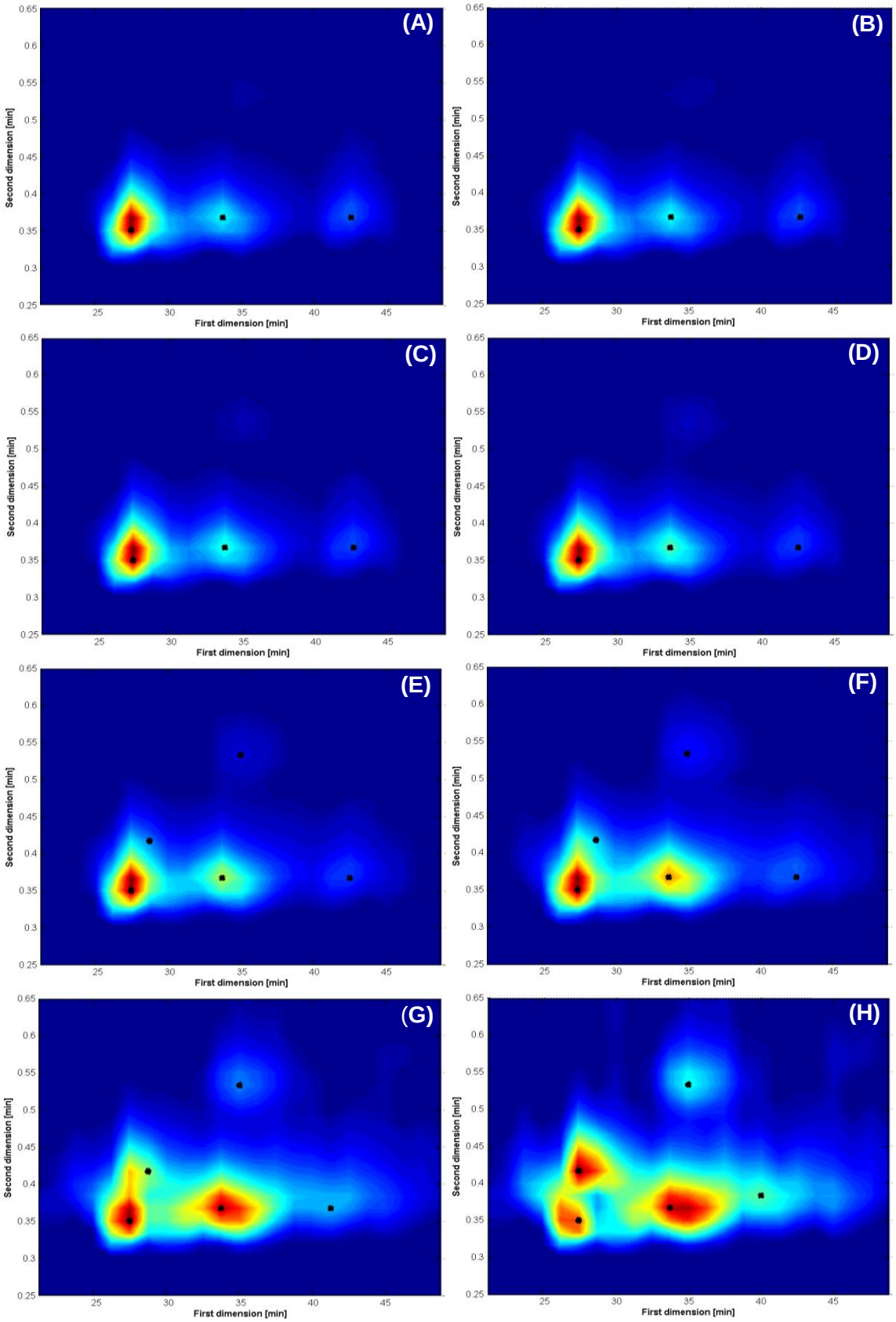

Fig. IV-6. Two-dimensional chromatographic sections of sample B at different wavelengths: 259 nm (**A**), 263 nm (**B**), 267 nm (**C**), 271 nm (**D**), 275 nm (**E**), 279 nm (**F**), 283 nm (**G**), and 287 nm (**H**).

The two-dimensional chromatogram of sample *A*, in Fig. IV-5A exhibit thirteen two-dimensional chromatographic peaks. The coordinates (i.e. retention times) of the two-dimensional peaks and their relative intensity are shown in Table IV-3. The most intense peak is peak **5**, immediately followed by peak **3**. All the remaining two-dimensional chromatographic peaks show a relatively low intensity because none of them reach 50% of the intensity of peak **5**.

Table IV-3. Two-dimensional peak location and relative intensity for sample *A*.

Peak	First dimension (min)	Second dimension (min)	Relative intensity (AU)
1	20.0	0.37	0.157
2	25.0	0.38	0.190
3	27.5	0.35	0.944
4	28.8	0.40	0.434
5	36.3	0.35	1.00
6	36.3	0.52	0.137
7	95.0	0.48	0.336
8	96.3	0.63	0.089
9	108	0.48	0.091
10	110	0.37	0.091
11	140	0.37	0.093
12	230	0.67	0.204
13	231	0.82	0.166

On the other hand, the two-dimensional chromatogram of sample *B*, shown in Fig. IV-5B, only presents eleven chromatographic peaks, whose features are described in Table IV-4. Peak **1** is the most intense peak, followed by peak **3**. Regarding the other two-dimensional chromatographic peaks, only two have an intensity higher than 30% of the intensity of peak **1**.

Table IV-4. Two-dimensional peak location and relative intensity for sample *B*.

Peak	First dimension (min)	Second dimension (min)	Relative intensity (AU)
1	27.5	0.350	1.00
2	28.8	0.417	0.361
3	33.8	0.367	0.788
4	35.0	0.533	0.152
5	42.5	0.367	0.220
6	93.8	0.483	0.168
7	103	0.483	0.178
8	123	0.367	0.156
9	125	0.367	0.182
10	223	0.683	0.182
11	223	0.850	0.138

As depicted in Fig. IV-5C, the two-dimensional chromatogram of sample *C* exhibits the same number of peaks of sample *B*. Peak **3** is the most intense peak, as shown in Table IV-5, followed by peak **1** with a relative intensity of 0.689. The intensity of the other two-dimensional peaks does not exceed 30% of that of peak **3**. It should also be mentioned that the two-dimensional chromatogram of sample *C* also presents a chromatographic artifact at about 190 min in the first dimension and all along the second dimension. The presence of this artifact can be justified by an unpredicted power oscillation episode during the chromatographic analysis.

Table IV-5. Two-dimensional peak location and relative intensity for sample *C*.

Peak	First dimension (min)	Second dimension (min)	Relative intensity (AU)
1	28.8	0.33	0.689
2	30.0	0.40	0.262
3	33.8	0.35	1.000
4	36.3	0.55	0.098
5	41.3	0.35	0.255
6	53.8	0.43	0.095
7	90.0	0.45	0.361
8	96.3	0.45	0.111
9	115	0.35	0.097
10	208	0.82	0.101
11	209	0.65	0.143

In general, and for all samples, the peaks can be considered as being distributed according to three main chromatographic regions: the first, roughly between 20 and 70 minutes in the first dimension and 0.3 and 0.6 min in the second dimension, the second between 85 and 150 min in the first dimension and 0.3 and 0.8 min in the second dimension, and the third between 200 and 240 min in the first dimension and 0.6 and 1.2 min in the second dimension. The first region encloses the largest number of two-dimensional peaks with the higher intensity. The second region is the region with more variability between samples and, consequently, constitutes the chromatographic region with more interest for wine fingerprinting. In contrast, the third region is almost similar for all chromatograms, only exhibiting a few variations, especially along the first dimension. Furthermore, it seems that there are some two-dimensional peaks which are common to all samples. For instance, peaks **3** and **5** in sample *A* and peaks **1** and **3** in samples *B* and C exhibit a similar chromatographic behaviour, though a small time shifts area observed between the samples. The same feature is verified for peaks **4** and **6** in sample *A* and peaks

2 and **4** in the other two samples. In spite of some retention times shifting along the second dimension, peaks **3** and **5** are common to both samples *B* and *C*. Nevertheless, it seems that the relative intensities these two peaks varies between the samples, which can be considered as an important factor for the fingerprinting of wine samples. On the other hand, although the other peaks appear to have the same chromatographic shape, the different retention times and intensities between the samples, hinders any valid conclusion. Despite of not having an enough number of samples to build a statistical support for the results, these analyses suggest the existence of some important similarities as well as variations between the samples and, consequently, a different fingerprint for each red wine.

4.3.2.3. Two-dimensional chromatographic profiles based on the fluorescence detector

The two-dimensional chromatograms of each sample obtained with the FLD at excitation/emission wavelengths of 230/320 nm are shown in Fig. IV-7. Since the signal-to-noise ratio is too high, the two-dimensional peak detection was not carried out. However, it is possible to perform a visual comparison between these chromatograms and those of the LC × LC-DAD in Fig. IV-5. All the three chromatograms represented in Fig. IV-7A show an intense fluorescence signal near the retention time of 100 min in the first dimension. When compared to the LC × LC-DAD chromatograms, this signal is located at the same retention time regions of peaks **7**, **8**, and **9** in Fig. IV-5A, of peaks **6** and **7** in Fig. IV-5B, and of peaks **7** and **8** in Fig. IV-5C. The LC × LC-FLD chromatograms of samples *B* and *C* exhibit another intense signal at the retention time of approximately 30 min in the first dimension. The LC × LC-DAD chromatograms of these two samples also exhibit a peak located at the same region, i.e., peak **2** (Tables IV-4 and IV-5, and Fig. IV-5B Fig. IV-5C)

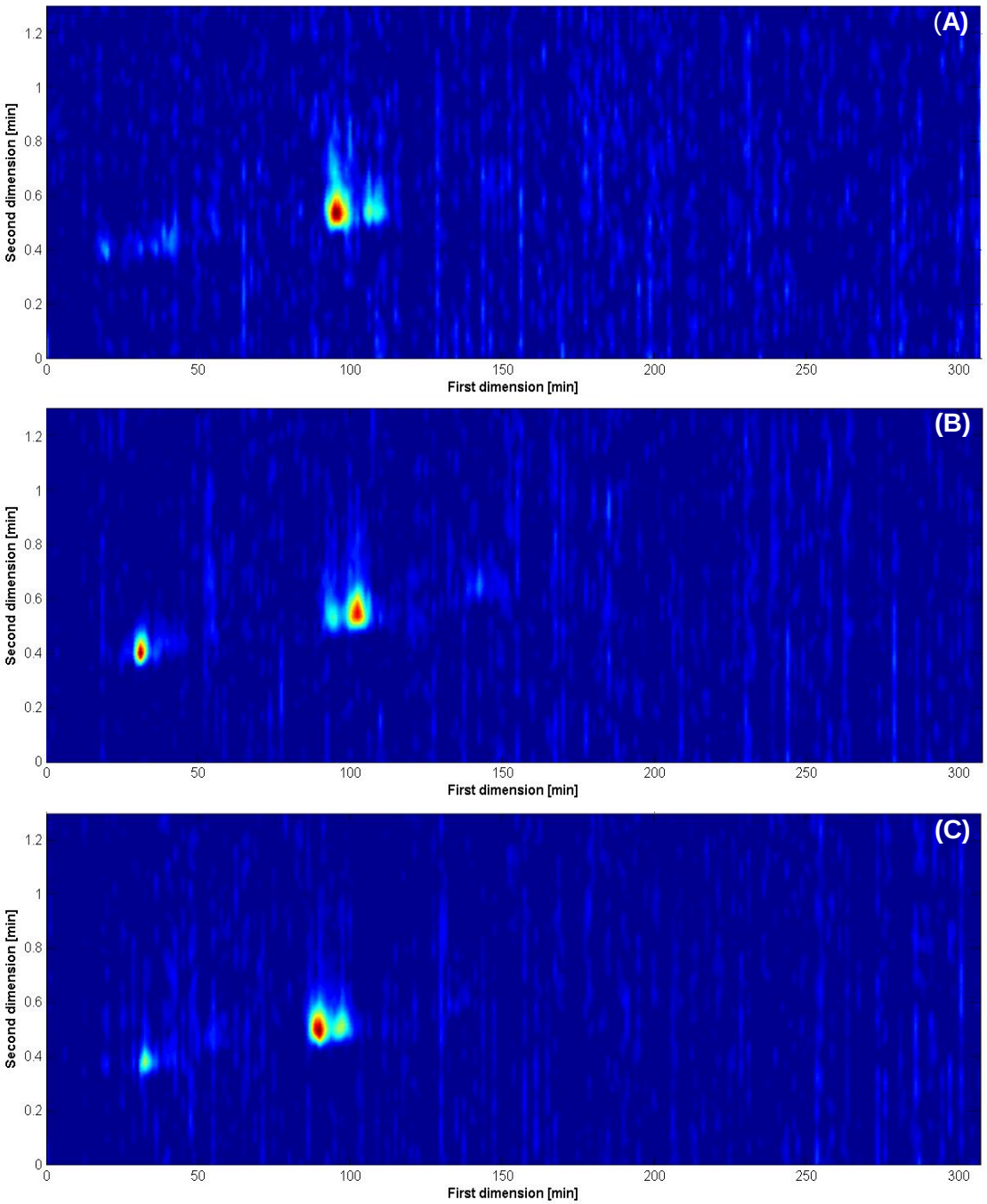

Fig. IV-7. Representation of the LC × LC-FLD at excitation/emission wavelengths of 230/320 nm for the wine sample *A* (**A**), wine sample *B* (**B**), and wine sample *C* (**C**).

4.3.2.4. Application of the two-dimensional chromatographic response function to red wine samples

With the aim of quantifying and compare each two-dimensional chromatographic red wine fingerprint, the new CRF_{2D} developed in the previous chapter was applied to the data set obtained with DAD at 279 nm. After data pre-treatment and the identification of the two-dimensional peaks, the results were fitted using a sum of two-dimensional peak functions (one for each overlapped peak) in order to find the best model that can describe the chromatographic behaviour of each sample. The mathematical peak model used in this study for quantifying the purity of the overlapped peaks is an adapted model that includes a three parameter Log Normal function in the first dimension and a Log-normal (Fraser-Suzuki) function in the second dimension, and it can be described as follows:

$$h_{(t_{1D},t_{2D})} = h_0 \exp\left\{-0.5\times\left\{\left[\frac{\ln\left(\frac{t_{1D}}{t_{R,1D}}\right)}{\sigma_{1D}}\right]^2\right\}\right\}\times\exp\left\{-\frac{\ln 2}{(\ln s_{2D})^2}\left\{\ln\left[\frac{(t_{2D}-t_{R,2D})(s_{2D}^2-1)}{\sigma_{2D}s_{2D}}+1\right]\right\}^2\right\} \quad \textbf{(45)}$$

where h_0 is the maximum peak height, σ_{1D} and σ_{2D} are the standard deviations along the first and second dimension, respectively, and s_{2D} is a parameter related to the symmetry of the peak and its coordinates ($t_{R,1D}$ and $t_{R,2D}$). In the fitting procedure, a Trust-Region Algorithm were used to assure the best result. The quality of the fitting is guaranteed by the relatively high R^2_{adj} values, ranging from 0.9155 to 0.9620 for all the overlapped cluster peaks in the chromatograms. Using the parameters estimated by this fitting procedure, the two-dimensional chromatographic peaks were simulated and the final results are shown in Fig. IV-8. For a better visualization and comparison of the results, well separated peaks were also added into these final representations. The final steps entailed the estimation of the purity of each overlapped peak and, afterwards, the computation of the CRF_{2D} for each wine sample. The results obtained for each sample are described in Table IV-6.

Table IV-6. Individual peak purity and CRF_{2D} values for the three red wine samples.

Sample	Peak purity (P_i) [(a)]													CRF_{2D} [(b)]
	1	2	3	4	5	6	7	8	9	10	11	12	13	
A	0.96	0.43	0.50	0.59	0.93	0.63	0.90	0.82	0.91	0.97	1.0	0.65	0.48	22.5
B	0.80	0.25	0.84	0.81	0.83	0.91	0.93	0.10	0.58	0.90	0.91	-	-	18.7
C	0.83	0.11	0.78	0.87	0.74	1.0	0.95	0.79	1.0	0.92	0.90	-	-	19.7

(a) Peak purity (P_i) calculated through Equation (40)).

(b) CRF_{2D} calculated through Equation (41) using the peak purity (P_i) concept of Equation (40).

Despite the fact that the number of studied samples are too low for providing a statistically support of the results, the CRF_{2D} values seem to be in accordance with chromatographic profile of the samples. Regarding the origin of the samples, and although they come from 3 different regions (section 4.2.1), huge differences in the values of the CFR_{2D} are not expected, mostly because the samples have their origin in relatively close regions and, in consequence, they are subjected to similar weather events. On the other hand, in what concerns the grapes used in the wine production, it is possible to verify that Sample *A* shares a grape with sample *B* (Aragonês) and two grapes with Sample C (Aragonês and Touriga Nacional). In addition, sample *B* also shares two grapes with sample *C* (Aragonês e Alfrocheiro). Therefore, it is expected that samples containing more than one similar type of grapes will exhibit close CRF_{2D} values. This assumption is supported by the experimental data, where the higher difference between the CRF_{2D} values is found for samples *A* and *B*; on the other hand, the differences between samples *A* and *C* and samples *B* and *C* are relatively lower.

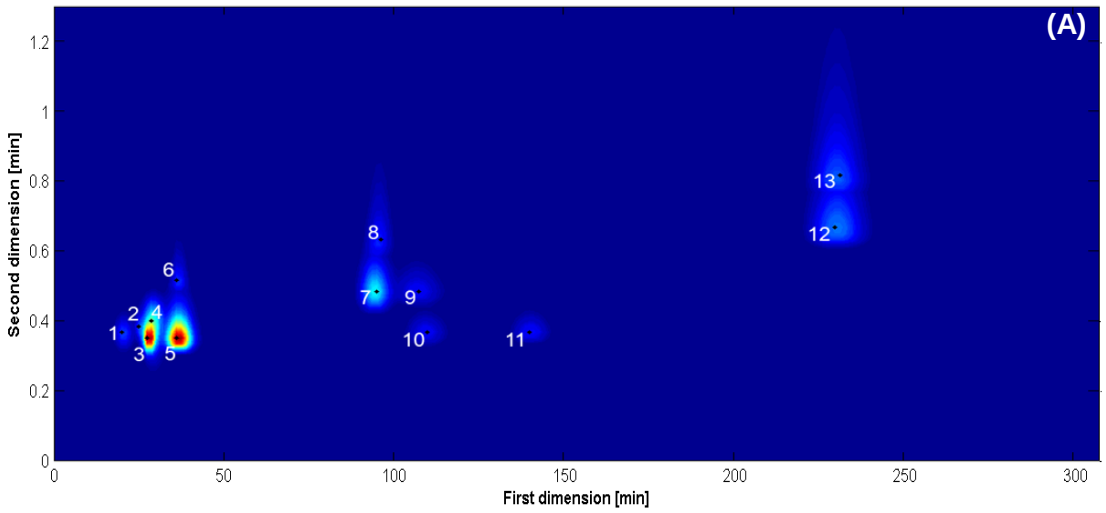

Fig. IV-8. Representation of the simulated chromatogram corresponding to the LC × LC-DAD chromatogram obtained at 279 nm for the wine sample *A* (**A**), wine sample *B* (**B**), and wine sample *C* (**C**).

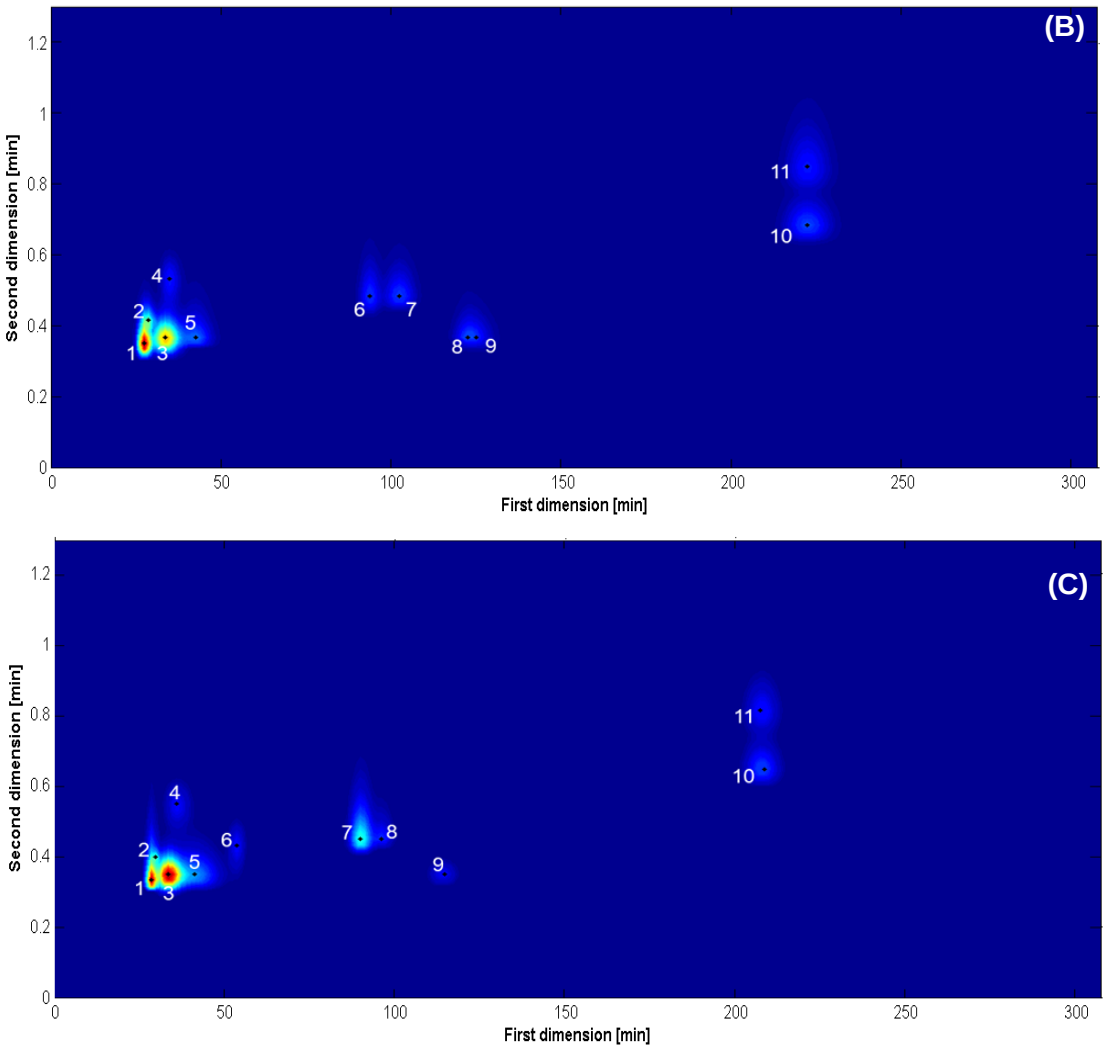

Fig.IV-8. Representation of the simulated chromatogram corresponding to the LC × LC-DAD chromatogram obtained at 279 nm for the wine sample *A* (**A**), wine sample *B* (**B**), and wine sample *C* (**C**). (cont.)

4.4. Conclusions and research needs

This work successfully applied a CRF_{2D} to three different two-dimensional liquid chromatograms in order to quantify the red wine fingerprinting. Nevertheless, there are some problems that need to be solved before a higher application of this methodology.

Firstly, it is necessary to develop better chromatographic conditions in order to solve the huge amount of sample information that come out in the void time. Although this problem can be partially solved in the pre-processing of data, it is very difficult to assure the inexistence of deformations in the two-dimensional peaks after the subtraction of the baseline. On the other hand, it is necessary to guarantee that the peak size is not incremented by the modulation valve signal. This can cause problems in the normalization of the data and can create problems for quantitative purposes. Another problem is the time spent in each chromatographic analysis. The major drawback that hinders the extensively implementation of this technique for wine fingerprinting is that a single injection took more than five hours to be accomplished. This means that it is necessary to study some chromatographic improvements in order to reduce the time of analysis. On the other hand, the data treatment technique that deals with all multichannel spectral information is still in early stages of development. The use of different DAD wavelengths for comparing the three wine samples is clearly unexploited, and much work still needs to be done in this field. A problem in the application of the CRF_{2D} for fingerprinting is related to the fact that this method does not take into account directly the peak intensity. This fact can be considered as a weakness since the different amounts of the same compounds in the samples are not being taken into account in the fingerprinting procedure.

V

Final remarks

This dissertation allowed the identification of some important issues that must be taken into account on the application of two-dimensional liquid chromatography to complex mixtures. Firstly, the critical review of the existing data processing methods allowed concluding on the lack of algorithms and tools to deal with some two-dimensional chromatographic problems. Although many aspects have already been addressed to deal with target analysis, processing higher-order data structures from non-target samples remains a problem in many aspects. Furthermore, in the most cases, the algorithms already developed are extremely complex, which makes impossible their wider application.

The new CRF_{2D} proposed in this dissertation was designed to deal with non-target two-dimensional chromatographic results. In the development of this CRF_{2D}, a two-dimensional version of the peak purity was presented and applied with success for measuring two-dimensional peaks overlap, thus producing better results than the already developed concepts. Furthermore, the new CRF_{2D} and the two-dimensional peak purity were only applied to bilinear data structures. Future work should expand these concepts for application in higher-order data structures.

Finally, this work presented a procedure for two-dimensional chromatographic fingerprinting of red wine samples, with promising results. However, much work is still needed for a wider application, starting by developing an enhanced two-dimensional chromatographic system in order to improve a better separation and reduce the time of analysis. On the other hand, the CRF_{2D} used for the fingerprinting quantification need to be also improved in order to take into account the peak intensity and all multichannel spectral information.

References

AIRADO-RODRÍGUEZ, D., GALEANO-DÍAZ, T., DURÁN-MERÁS, I. & WOLD, J. P. 2009. Usefulness of Fluorescence Excitation−Emission Matrices in Combination with PARAFAC, as Fingerprints of Red Wines. *Journal of Agricultural and Food Chemistry,* 57**,** 1711-1720.

ALLEN, R. C. & RUTAN, S. C. 2011. Investigation of interpolation techniques for the reconstruction of the first dimension of comprehensive two-dimensional liquid chromatography–diode array detector data. *Analytica Chimica Acta,* 705**,** 253-260.

AMIGO, J. M., SKOV, T. & BRO, R. 2010. ChroMATHography: Solving Chromatographic Issues with Mathematical Models and Intuitive Graphics. *Chemical Reviews,* 110**,** 4582-4605.

ANDERSSON, C. A. & BRO, R. 2000. The N-way Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems,* 52**,** 1-4.

ARVANITOYANNIS, I. S., KATSOTA, M. N., PSARRA, E. P., SOUFLEROS, E. H. & KALLITHRAKA, S. 1999. Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends in Food Science & Technology,* 10**,** 321-336.

BAILEY, H. P. & RUTAN, S. C. 2011. Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine. *Chemometrics and Intelligent Laboratory Systems,* 106**,** 131-141.

BAILEY, H. P., RUTAN, S. C. & CARR, P. W. 2011. Factors that affect quantification of diode array data in comprehensive two-dimensional liquid chromatography using chemometric data analysis. *Journal of Chromatography A.*

BERRIDGE, J. C. 1982. Unattended optimisation of reversed-phase high-performance liquid chromatographic separations using the modified simplex algorithm. *Journal of Chromatography A,* 244**,** 1-14.

BOLANČA, T. & CERJAN-STEFANOVIĆ, Š. 2007. Optimization Strategies in Ion Chromatography. *Journal of Liquid Chromatography & Related Technologies,* 30**,** 791-806.

BOSTYN, S., CAGNON, B. & FAUDUET, H. 2009. Optimization by the simplex method of the separation of phenolic acids by high-performance liquid chromatography in wastewater olive and sugar beet vinasse. *Talanta,* 80**,** 1-7.

BRO, R. 1997. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems,* 38**,** 149-171.

BRUCKNER, C. A., PRAZEN, B. J. & SYNOVEC, R. E. 1998. Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis. *Analytical Chemistry,* 70**,** 2796-2804.

BYLUND, D., BERGENS, A. & JACOBSSON, S. 1997. Optimisation of chromatographic separations by use of a chromatographic response function, empirical modelling and multivariate analysis. *Chromatographia,* 44**,** 74-80.

CACCIOLA, F., JANDERA, P., HAJDÚ, Z., ČESLA, P. & MONDELLO, L. 2007. Comprehensive two-dimensional liquid chromatography with parallel gradients for

separation of phenolic and flavone antioxidants. *Journal of Chromatography A,* 1149**,** 73-87.

CARLE, G. C. 1972. Determination of chromatographic resolution for peaks of vast concentration differences. *Analytical Chemistry,* 44**,** 1905-1906.

CHIRA, K., PACELLA, N., JOURDES, M. & TEISSEDRE, P.-L. 2011. Chemical and sensory evaluation of Bordeaux wines (Cabernet-Sauvignon and Merlot) and correlation with wine age. *Food Chemistry,* 126**,** 1971-1977.

CORDERO, C., LIBERTO, E., BICCHI, C., RUBIOLO, P., REICHENBACH, S. E., TIAN, X. & TAO, Q. 2010. Targeted and non-targeted approaches for complex natural sample profiling by GC×GC-qMS. *Journal of Chromatographic Science,* 48**,** 251-261.

DAVIS, J. M. 1991. Statistical theory of spot overlap in two-dimensional separations. *Analytical Chemistry,* 63**,** 2141-2152.

DI MARCO, V. B. & BOMBI, G. G. 2001. Mathematical functions for the representation of chromatographic peaks. *Journal of Chromatography A,* 931**,** 1-30.

DUARTE, R. M. B. O. & DUARTE, A. C. 2010. A new chromatographic response function for use in size-exclusion chromatography optimization strategies: Application to complex organic mixtures. *Journal of Chromatography A,* 1217**,** 7556-7563.

DUGO, P., CACCIOLA, F., DONATO, P., AIRADO-RODRÍGUEZ, D., HERRERO, M. & MONDELLO, L. 2009. Comprehensive two-dimensional liquid chromatography to quantify polyphenols in red wines. *Journal of Chromatography A,* 1216**,** 7483-7487.

ERNI, F. & FREI, R. W. 1978. Two-dimensional column liquid chromatographic technique for resolution of complex mixtures. *Journal of Chromatography A,* 149**,** 561-569.

FALCÃO, L. D., DE REVEL, G., ROSIER, J. P. & BORDIGNON-LUIZ, M. T. 2008. Aroma impact components of Brazilian Cabernet Sauvignon wines using detection frequency analysis (GC–olfactometry). *Food Chemistry,* 107**,** 497-505.

FRAGA, C. G., BRUCKNER, C. A. & SYNOVEC, R. E. 2000a. Increasing the Number of Analyzable Peaks in Comprehensive Two-Dimensional Separations through Chemometrics. *Analytical Chemistry,* 73**,** 675-683.

FRAGA, C. G., PRAZEN, B. J. & SYNOVEC, R. E. 2000b. Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the High-Speed Quantitative Analysis of Aromatic Isomers in a Jet Fuel Using the Standard Addition Method and an Objective Retention Time Alignment Algorithm. *Analytical Chemistry,* 72**,** 4154-4162.

FRAGA, C. G., PRAZEN, B. J. & SYNOVEC, R. E. 2001. Objective Data Alignment and Chemometric Analysis of Comprehensive Two-Dimensional Separations with Run-to-Run Peak Shifting on Both Dimensions. *Analytical Chemistry,* 73**,** 5833-5840.

FRAGA, C. G. & CORLEY, C. A. 2005. The chemometric resolution and quantification of overlapped peaks form comprehensive two-dimensional liquid chromatography. *Journal of Chromatography A,* 1096**,** 40-49.

GARCÍA-ÁLVAREZ-COQUE, M. C., TORRES-LAPASIÓ, J. R. & BAEZA-BAEZA, J. J. 2006. Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography. *Analytica Chimica Acta,* 579**,** 125-145.

GHESHLAGHI, R., SCHARER, J. M., MOO-YOUNG, M. & DOUGLAS, P. L. 2008. Application of statistical design for the optimization of amino acid separation by reverse-phase HPLC. *Analytical Biochemistry,* 383**,** 93-102.

GIDDINGS, J. C. 1990. *Multidimensional Chromatography: techniques and applications,* New York, NY.

GÓMEZ-ALONSO, S., GARCÍA-ROMERO, E. & HERMOSÍN-GUTIÉRREZ, I. 2007. HPLC analysis of diverse grape and wine phenolics using direct injection and multidetection by DAD and fluorescence. *Journal of Food Composition and Analysis,* 20**,** 618-626.

GUADALUPE, Z., MARTÍNEZ-PINILLA, O., GARRIDO, Á., CARRILLO, J. D. & AYESTARÁN, B. 2012. Quantitative determination of wine polysaccharides by gas chromatography–mass spectrometry (GC–MS) and size exclusion chromatography (SEC). *Food Chemistry,* 131**,** 367-374.

GUIOCHON, G., MARCHETTI, N., MRIZIQ, K. & SHALLIKER, R. A. 2008. Implementations of two-dimensional liquid chromatography. *Journal of Chromatography A,* 1189**,** 109-168.

HENDERSON, J. P. & REX, D. 2011. *About Wine,* New York, Cengage Learning.

HOGGARD, J. C. & SYNOVEC, R. E. 2007. Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors. *Analytical Chemistry,* 79**,** 1611-1619.

HOGGARD, J. C. & SYNOVEC, R. E. 2008. Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis. *Analytical Chemistry,* 80**,** 6677-6688.

HOLLINGSWORTH, B. V., REICHENBACH, S. E., TAO, Q. & VISVANATHAN, A. 2006. Comparative visualization for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A,* 1105**,** 51-58.

HOPE, J. L., SINHA, A. E., PRAZEN, B. J. & SYNOVEC, R. E. 2005. Evaluation of the DotMap algorithm for locating analytes of interest based on mass spectral similarity in data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry. *Journal of Chromatography A,* 1086**,** 185-192.

JOHNSON, K. J., PRAZEN, B. J., YOUNG, D. C. & SYNOVEC, R. E. 2004. Quantification of naphthalenes in jet fuel with GC×GC/Tri-PLS and windowed rank minimization retention time alignment. *Journal of Separation Science,* 27**,** 410-416.

KIVILOMPOLO, M. & HYÖTYLÄINEN, T. 2007. Comprehensive two-dimensional liquid chromatography in analysis of Lamiaceae herbs: Characterisation and quantification of antioxidant phenolic acids. *Journal of Chromatography A,* 1145**,** 155-164.

KIVILOMPOLO, M., OBŮRKA, V. & HYÖTYLÄINEN, T. 2008. Comprehensive two-dimensional liquid chromatography in the analysis of antioxidant phenolic compounds in wines and juices. *Analytical and Bioanalytical Chemistry,* 391**,** 373-380.

LATHA, I., REICHENBACH, S. E. & TAO, Q. 2011. Comparative analysis of peak-detection techniques for comprehensive two-dimensional chromatography. *Journal of Chromatography A,* In Press, Corrected Proof.

LOHNES, M. T., GUY, R. D. & WENTZELL, P. D. 1999. Window target-testing factor analysis: theory and application to the chromatographic analysis of complex mixtures with multiwavelength fluorescence detection. *Analytica Chimica Acta,* 389**,** 95-113.

MARCOBAL, A., POLO, M. C., MARTíN-ÁLVAREZ, P. J. & MORENO-ARRIBAS, M. V. 2005. Biogenic amine content of red Spanish wines: comparison of a direct ELISA and an HPLC method for the determination of histamine in wines. *Food Research International,* 38**,** 387-394.

MICYUS, N. J., SEELEY, S. K. & SEELEY, J. V. 2005. Method for reducing the ambiguity of comprehensive two-dimensional chromatography retention times. *Journal of Chromatography A,* 1086**,** 171-174.

MO DUGO, G., VILASI, F., LA TORRE, G. L. & PELLICANÒ, T. M. 2006. Reverse phase HPLC/DAD determination of biogenic amines as dansyl derivatives in experimental red wines. *Food Chemistry,* 95**,** 672-676.

MORET, I., SCARPONI, G. & CESCON, P. 1994. Chemometric Characterization and Classification of Five Venetian White Wines. *Journal of Agricultural and Food Chemistry,* 42**,** 1143-1153.

MORRIS, V. M., HUGHES, J. G. & MARRIOTT, P. J. 1996. Examination of a new chromatographic function, based on an exponential resolution term, for use in optimization strategies: application to capillary gas chromatography separation of phenols. *Journal of Chromatography A,* 755**,** 235-243.

NELSON, R. K., KILE, B. M., PLATA, D. L., SYLVA, S. P., XU, L., REDDY, C. M., GAINES, R. B., FRYSINGER, G. S. & REICHENBACH, S. E. 2006. Tracking the Weathering of an Oil Spill with Comprehensive Two-Dimensional Gas Chromatography. *Environmental Forensics,* 7**,** 33-44.

NIKITAS, P. & PAPPA-LOUISI, A. 2009. Retention models for isocratic and gradient elution in reversed-phase liquid chromatography. *Journal of Chromatography A,* 1216**,** 1737-1755.

ORTÍN, A., TORRES-LAPASIÓ, J. R. & GARCÍA-ÁLVAREZ-COQUE, M. C. 2011. Finding the best separation in situations of extremely low chromatographic resolution. *Journal of Chromatography A,* 1218**,** 2240-2251.

PETERS, S., VIVÓ-TRUYOLS, G., MARRIOTT, P. J. & SCHOENMAKERS, P. J. 2007a. Development of an algorithm for peak detection in comprehensive two-dimensional chromatography. *Journal of Chromatography A,* 1156**,** 14-24.

PETERS, S., VIVÓ-TRUYOLS, G., MARRIOTT, P. J. & SCHOENMAKERS, P. J. 2007b. Development of a resolution metric for comprehensive two-dimensional chromatography. *Journal of Chromatography A,* 1146**,** 232-241.

PHILLIPS, J. B. & BEENS, J. 1999. Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions. *Journal of Chromatography A,* 856**,** 331-347.

PIERCE, K. M., WOOD, L. F., WRIGHT, B. W. & SYNOVEC, R. E. 2005. A Comprehensive Two-Dimensional Retention Time Alignment Algorithm To Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data. *Analytical Chemistry,* 77**,** 7735-7743.

PIERCE, K. M., HOGGARD, J. C., MOHLER, R. E. & SYNOVEC, R. E. 2008. Recent advancements in comprehensive two-dimensional separations with chemometrics. *Journal of Chromatography A,* 1184**,** 341-352.

PÓL, J., HOHNOVÁ, B., JUSSILA, M. & HYÖTYLÄINEN, T. 2006. Comprehensive two-dimensional liquid chromatography–time-of-flight mass spectrometry in the analysis of acidic compounds in atmospheric aerosols. *Journal of Chromatography A,* 1130**,** 64-71.

PORTER, S. E. G., STOLL, D. R., RUTAN, S. C., CARR, P. W. & COHEN, J. D. 2006. Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics. *Analytical Chemistry,* 78**,** 5559-5569.

PRAZEN, B. J., SYNOVEC, R. E. & KOWALSKI, B. R. 1998. Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis. *Analytical Chemistry,* 70**,** 218-225.

PRAZEN, B. J., BRUCKNER, C. A., SYNOVEC, R. E. & KOWALSKI, B. R. 1999a. Second-order chemometric standardization for high-speed hyphenated gas chromatography: Analysis of GC/MS and comprehensive GC×GC data. *Journal of Microcolumn Separations,* 11**,** 97-107.

PRAZEN, B. J., BRUCKNER, C. A., SYNOVEC, R. E. & KOWALSKI, B. R. 1999b. Enhanced Chemical Analysis Using Parallel Column Gas Chromatography with Single-Detector Time-of-Flight Mass Spectrometry and Chemometric Analysis. *Analytical Chemistry,* 71**,** 1093-1099.

REICHENBACH, S. E., NI, M., ZHANG, D. & LEDFORD, E. B. 2003. Image background removal in comprehensive two-dimensional gas chromatography. *Journal of Chromatography A,* 985**,** 47-56.

REICHENBACH, S. E., NI, M., KOTTAPALLI, V. & VISVANATHAN, A. 2004. Information technologies for comprehensive two-dimensional gas chromatography. *Chemometrics and Intelligent Laboratory Systems,* 71**,** 107-120.

REICHENBACH, S. E., KOTTAPALLI, V., NI, M. & VISVANATHAN, A. 2005. Computer language for identifying chemicals with comprehensive two-dimensional gas chromatography and mass spectrometry. *Journal of Chromatography A,* 1071**,** 263-269.

REICHENBACH, S. E., CARR, P. W., STOLL, D. R. & TAO, Q. 2009. Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography. *Journal of Chromatography A,* 1216**,** 3458-3466.

REICHENBACH, S. E., TIAN, X., CORDERO, C. & TAO, Q. 2011. Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography. *Journal of Chromatography A*.

RODRIGUES, S. M., OTERO, M., ALVES, A. A., COIMBRA, J., COIMBRA, M. A., PEREIRA, E. & DUARTE, A. C. 2009. Elemental analysis for categorization of wines and

authentication of their certified brand of origin. *Journal of Food Composition and Analysis,* 24**,** 548-562.

SANCHEZ, E. & KOWALSKI, B. R. 1986. Generalized rank annihilation factor analysis. *Analytical Chemistry,* 58**,** 496-499.

SANTOS, J. P., ARROYO, T., ALEIXANDRE, M., LOZANO, J., SAYAGO, I., GARCíA, M., FERNÁNDEZ, M. J., ARÉS, L., GUTIÉRREZ, J., CABELLOS, J. M., GIL, M. & HORRILLO, M. C. 2004. A comparative study of sensor array and GC–MS: application to Madrid wines characterization. *Sensors and Actuators B: Chemical,* 102**,** 299-307.

SAVITZKY, A. & GOLAY, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry,* 36**,** 1627-1639.

SCHURE, M. R. 1997. Quantification of resolution for two-dimensional separations. *Journal of Microcolumn Separations,* 9**,** 169-176.

ŠERUGA, M., NOVAK, I. & JAKOBEK, L. 2011. Determination of polyphenols content and antioxidant activity of some red wines by differential pulse voltammetry, HPLC and spectrophotometric methods. *Food Chemistry,* 124**,** 1208-1216.

SHI, W. & DAVIS, J. M. 1993. Test of theory of overlap for two-dimensional separations by computer simulations of three-dimensional concentration profiles. *Analytical Chemistry,* 65**,** 482-492.

SINHA, A. E., JOHNSON, K. J., PRAZEN, B. J., LUCAS, S. V., FRAGA, C. G. & SYNOVEC, R. E. 2003. Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument. *Journal of Chromatography A,* 983**,** 195-204.

SINHA, A. E., FRAGA, C. G., PRAZEN, B. J. & SYNOVEC, R. E. 2004a. Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography–time-of-flight mass spectrometry data. *Journal of Chromatography A,* 1027**,** 269-277.

SINHA, A. E., HOPE, J. L., PRAZEN, B. J., FRAGA, C. G., NILSSON, E. J. & SYNOVEC, R. E. 2004b. Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry subjected to chemometric peak deconvolution. *Journal of Chromatography A,* 1056**,** 145-154.

SINHA, A. E., HOPE, J. L., PRAZEN, B. J., NILSSON, E. J., JACK, R. M. & SYNOVEC, R. E. 2004c. Algorithm for locating analytes of interest based on mass spectral similarity in GC × GC-TOF-MS data: analysis of metabolites in human infant urine. *Journal of Chromatography A,* 1058**,** 209-215.

SIOUFFI, A. M. & PHAN-TAN-LUU, R. 2000. Optimization methods in chromatography and capillary electrophoresis. *Journal of Chromatography A,* 892**,** 75-106.

STOLL, D. R., COHEN, J. D. & CARR, P. W. 2006. Fast, comprehensive online two-dimensional high performance liquid chromatography through the use of high temperature ultra-fast gradient elution reversed-phase liquid chromatography. *Journal of Chromatography A,* 1122**,** 123-137.

STOLL, D. R., LI, X., WANG, X., CARR, P. W., PORTER, S. E. G. & RUTAN, S. C. 2007. Fast, comprehensive two-dimensional liquid chromatography. *Journal of Chromatography A,* 1168**,** 3-43.

TARANTILIS, P. A., TROIANOU, V. E., PAPPAS, C. S., KOTSERIDIS, Y. S. & POLISSIOU, M. G. 2008. Differentiation of Greek red wines on the basis of grape variety using attenuated total reflectance Fourier transform infrared spectroscopy. *Food Chemistry,* 111**,** 192-196.

TOMASI, G. & BRO, R. 2006. A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis,* 50**,** 1700-1734.

VAN MISPELAAR, V. G., TAS, A. C., SMILDE, A. K., SCHOENMAKERS, P. J. & VAN ASTEN, A. C. 2003. Quantitative analysis of target components by comprehensive two-dimensional gas chromatography. *Journal of Chromatography A,* 1019**,** 15-29.

VIVÓ-TRUYOLS, G. & JANSSEN, H.-G. 2010. Probability of failure of the watershed algorithm for peak detection in comprehensive two-dimensional chromatography. *Journal of Chromatography A,* 1217**,** 1375-1385.

WARDLAW, G. D., AREY, J. S., REDDY, C. M., NELSON, R. K., VENTURA, G. T. & VALENTINE, D. L. 2008. Disentangling Oil Weathering at a Marine Seep Using GC×GC: Broad Metabolic Specificity Accompanies Subsurface Petroleum Biodegradation. *Environmental Science & Technology,* 42**,** 7166-7173.

WELDEGERGIS, B. T., VILLIERS, A. D., MCNEISH, C., SEETHAPATHY, S., MOSTAFA, A., GÓRECKI, T. & CROUCH, A. M. 2011. Characterisation of volatile components of Pinotage wines using comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC × GC–TOFMS). *Food Chemistry,* 129**,** 188-199.

XIE, L., MARRIOTT, P. J. & ADAMS, M. 2003. Chemometric analysis of comprehensive two-dimensional gas chromatography data using cryogenic modulation. *Analytica Chimica Acta,* 500**,** 211-222.

YIN, C., LI, H., DING, C. & WANG, H. 2009. Preliminary Investigation on Variety, Brewery and Vintage of Wines using Three-dimensional Fluorescence Spectroscopy. *Food Science and Technology Research,* 15**,** 27-38.

ZENG, Z.-D., CHIN, S.-T., HUGEL, H. M. & MARRIOTT, P. J. 2011. Simultaneous deconvolution and re-construction of primary and secondary overlapping peak clusters in comprehensive two-dimensional gas chromatography. *Journal of Chromatography A,* 1218**,** 2301-2310.

ZHANG, D., HUANG, X., REGNIER, F. E. & ZHANG, M. 2008. Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC−MS Data. *Analytical Chemistry,* 80**,** 2664-2671.

ZHANG, Y., WU, H.-L., XIA, A. L., HU, L.-H., ZOU, H.-F. & YU, R.-Q. 2007. Trilinear decomposition method applied to removal of three-dimensional background drift in comprehensive two-dimensional separation data. *Journal of Chromatography A,* 1167**,** 178-183.

Annexes

Description of the wine samples

Table A1. Sample *A* description.

Name:	Vale de São Domingos (Reserva)
Vintage:	2008
Grapes:	Tinta Roriz, Touriga Nacional, Touriga Franca, Tinta Barroca
Origen region:	Douro
Winemaker:	Castelinho Vinhos S.A. São João da Panasqueira – Douro - Portugal
E-mail:	castelinho@castelinho.vinhos.pt
Website:	www.castelinho-vinhos.com
Wine identification:	5 603196 003135

Table A2. Sample *B* description.

Name:	Morgado Do Reguengo
Vintage:	2008
Grapes:	Aragonês, Alfrocheiro e Castelões
Origen region:	Alentejo
Winemaker:	José Carvalho – Sociedade Agrícola Lda. Herdade dos Muachos 7300-575 Urra Portugal
E-mail:	herdademuachos@gmail.com
Website:	www.muachos.com
Wine identification:	5 609595 900181

Table A3. Sample *C* description.

Name:	Monte do Amante "Private Collection"
Vintage:	2008
Grapes:	Touriga Nacional, Alfrocheiro, Tinta-Roriz, and Jaen
Origen region:	Dão
Winemaker:	Vinícola de Nelas, SA-3520-061 Nelas, Portugal (em exclusivo para Luís Portugal Vinhos, LDA)
E-mail:	-
Website:	www.vinicola-nelas.pt
Wine identification:	5 601852 002072

Chromatograms of sample *A*

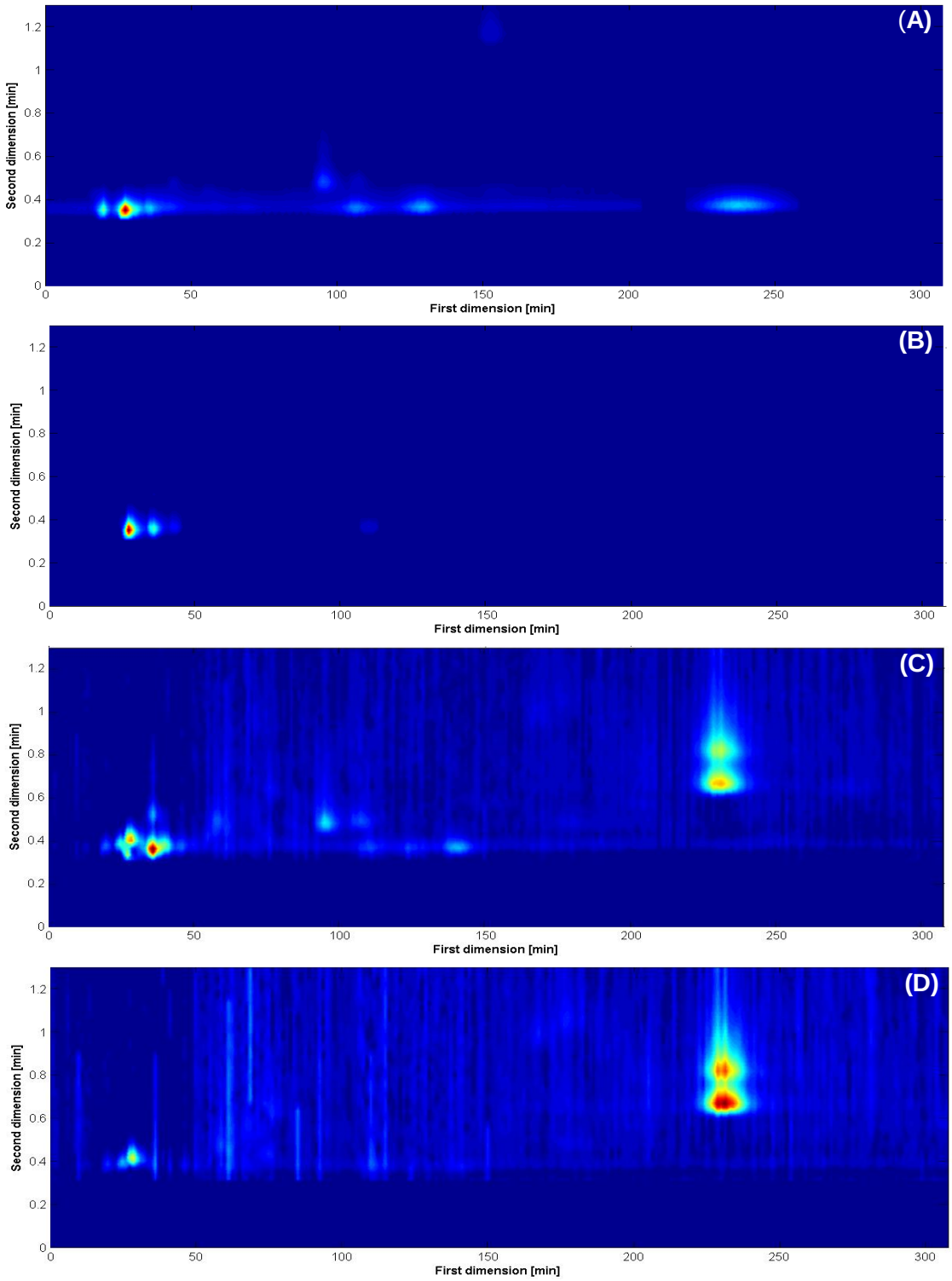

Fig. A1. Representation of the LC × LC-DAD of the wine sample *A* for the 215nm (**A**), 259nm (**B**), 287nm (**C**), and 311nm (**D**) wavelengths.

Chromatograms of sample *B*

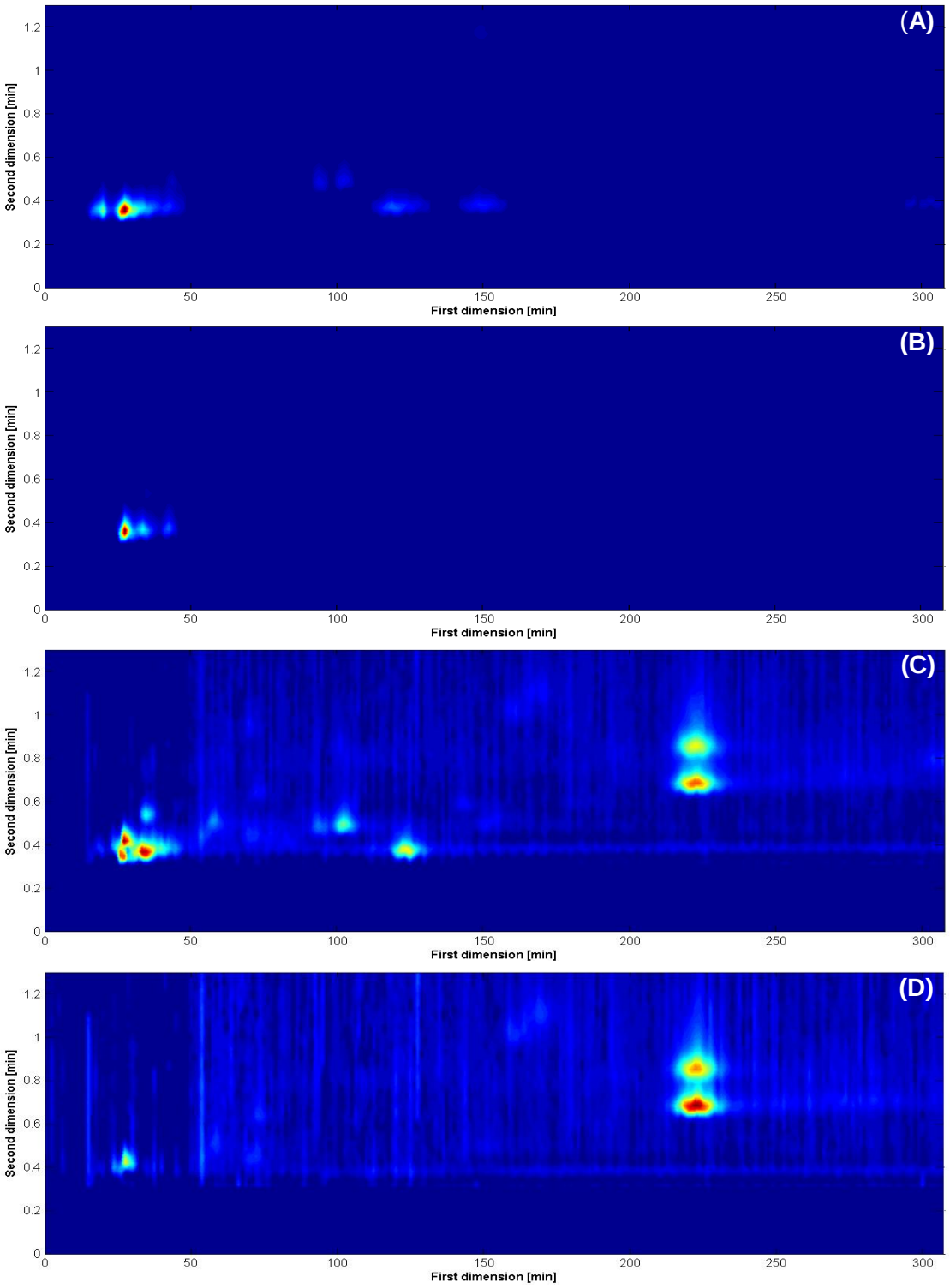

Fig. A2. Representation of the LC × LC-DAD of the wine sample *B* for the 215nm (**A**), 259nm (**B**), 287nm (**C**), and 311nm (**D**) wavelengths.

Chromatograms of sample *C*

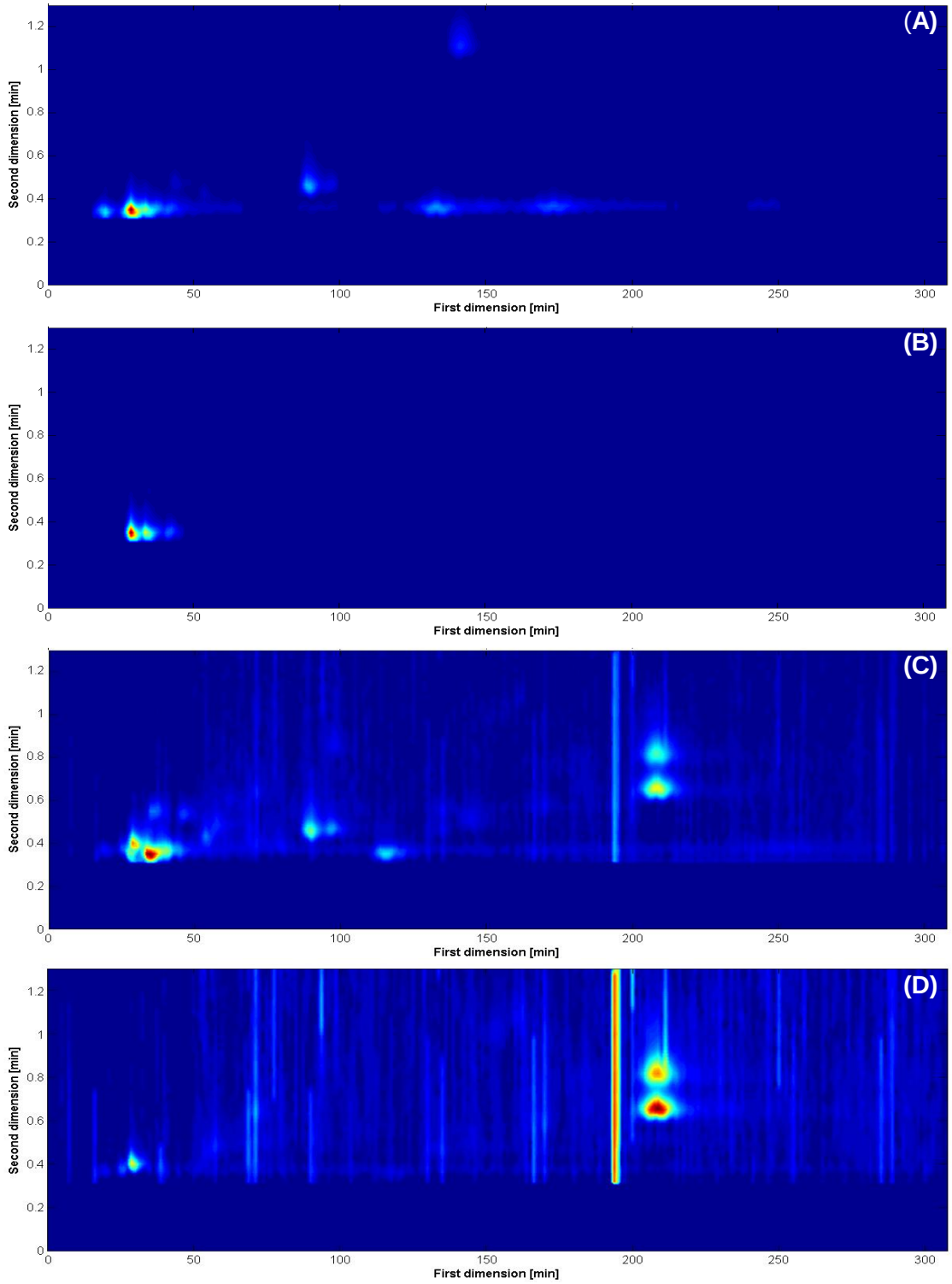

Fig. A3. Representation of the LC × LC-DAD of the wine sample *C* for the 215nm (**A**), 259nm (**B**), 287nm (**C**), and 311nm (**D**) wavelengths.